The importance to manage data protection in the right way: problems and solutions

Ph.D thesis in

“Computer science: Risk Management and Cybersecurity”

By: Hassan Mokalled

Industrial PhD at Hitachi Rail STS Company by:

Hassan Mokalled

In collaboration with:

UNIVERSITÀ DEGLI STUDI
DI GENOVA

Ph.D thesis in

“Computer science: Risk Management and Cybersecurity”

Supervised by:

Dr. Ermete Meda (Hitachi Rail STS Company)

Prof. Rodolfo Zunino (University of Genoa)

Prof. Abbas Hijazi and Prof. Ali Jaber (Lebanese University)

"Your work is going to fill a large part of your life, and the only way to be truly satisfied is to do what you believe is great work. And the only way to do great work is to love what you do. If you haven't found it yet, keep looking. Don't settle. As with all matters of the heart, you'll know when you find it."

Steve Jobs

Abstract

Information and communication technology (ICT) has made remarkable impact on the society, especially on companies and organizations. The use of computers, databases, servers, and other technologies has made an evolution on the way of storing, processing, and transferring data. However, companies access and share their data on internet or intranet, thus there is a critical need to protect this data from destructive forces and from the unwanted actions of unauthorized users. This thesis groups a set of solutions proposed, from a company point of view, to reach the goal of "Managing data protection". The work presented in this thesis represents a set of security solutions, which focuses on the management of data protection taking into account both the organizational and technological side. The work achieved can be divided into set of goals that are obtained particularly from the needs of the research community. This thesis handles the issue of managing data protection in a systematic way, through proposing a Data protection management approach, aiming to protect the data from both the organizational and the technological side, which was inspired by the ISO 27001 requirements. An Information Security Management System (ISMS) is then presented implementing this approach, an ISMS consists of the policies, procedures, guidelines, and associated resources and activities, collectively managed by an organization, in the pursuit of protecting its information assets. An ISMS is a systematic approach for establishing, implementing, operating, monitoring, reviewing, maintaining and improving an organization's information security to achieve business objectives, The goal of ISMS is to minimize risk and ensure continuity by pro-actively limiting the impact of a security breach. To be well-prepared to the potential threats that could occur to an organization, it is important to adopt an ISMS that helps in managing the data protection process, and in saving time and effort, minimizes cost of any loss. After that, a comprehensive framework is designed for the security risk management of Cyber Physical Systems (CPSs), this framework represents the strategy used to manage the security risk management, and it falls inside the ISMS as a security strategy. Traditional IT risk assessment methods can do the job (security risk management for a CPS); however, and because of the characteristics of a CPS, it is more efficient to adopt a solution that is wider than a method that addresses the type, functionalities and complexity of a CPS. Therefore, there is a critical need to follow a solution that breaks the restriction to a traditional risk assessment method, and so a high-level framework is proposed, it encompasses wider set of procedures and gives a great attention to the cybersecurity of these systems, which consequently leads to the safety of the physical world. In addition, inside the ISMS, another part of the work takes place, suggesting the guidelines to select an applicable Security Incident and Event Management (SIEM) solution. It also proposes an approach

that aims to support companies seeking to adopt SIEM systems into their environments, suggesting suitable answers to preferred requirements that are believed to be valuable prerequisites a SIEM system should have; and to suggest criteria to judge SIEM systems using an evaluation process composed of quantitative and qualitative methods. This approach, unlike others, is customer driven which means that customer needs are taken into account when following the whole approach, specifically when defining the requirements and then evaluating the suppliers' solutions. At the end, a research activity was carried out aiming classify web attacks on the network level, since any information about the attackers might be helpful and worth a lot to the cyber security analysts. And so, using network statistical fingerprints and machine learning techniques, a two-layers classification system is designed to detect the type of the web attack and the type of software used by the attackers.

Keywords: *Risk Management, cybersecurity, Information Security Management System, data protection, cyber physical systems, Security Information and Event Management, Intrusion detection, web attacks, Machine learning.*

Dedication

I am entirely grateful to the person who taught me the value of hard work in this life, to my father who passed away few months before I started my Ph.D. His absence did not stopped me; however, he gave me the persistence to continue what he started. He dedicated his entire life for his family, so I am dedicating this thesis for him even it represents nothing in return. Thank you so much "Dad", I will never forget you, your soul is always guiding me even you are not present.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor at HITACHI Rail STS Company Dr. Ermete Meda, with his immense knowledge; he guided me in both my research and professional work at the Company; also for his invaluable suggestions, contributions and motivation. He was always beside me during my Ph.D work in the last three years. I really cannot express my thanks and gratitude in just few lines. I might only say that I could not have imagined having better advisor for my Ph.D study than him.

Secondly, my deepest thankfulness to my advisors at the University, Prof. Rodolfo Zunino (University of Genoa), Prof. Abbas Hijazi (Lebanese University) and Prof. Ali Jaber (Lebanese University), I am very thankful for their continuous support in my Ph.D study and related research. With their deep experience and knowledge, they have guided me carefully to advance in my PhD.

Beside my advisors, I would like to thank the responsibles of our Ph.D cycle Prof. Mario Marchese (University of Genoa) and Prof. Hussein Chible (Lebanese University) for their continuous support, guidance, and for their incredible efforts, not just in my Ph.D work, but also facilitating many stages in my life during the last three years, both in Italy and Lebanon.

I would also like to thank my colleagues in the Cyber Security Assurance and Control department at HITACHI Rail STS Company. I start by our responsible Tina PRAGLIOLA, thanks a lot for her guidance, especially advising me how to balance between the research and professional work, and finding the best way to publish my work. Special thanks to Giovanna CENERE for her help, with her usual cheer and joy, giving me the motivation even in the toughest conditions far away from my family. Daniele DEBERTOL, who I would like to call

him my Ph.D “Unofficial” supervisor, always knowing what needs to be done in the best way. Even though Tina and Daniele were not officially my supervisors, I would like to thank them from the bottom of my heart for the time that they dedicated for me throughout the years of my PhD. At the end, Maximilian MORIELLA and Massimo MALAGMABA, although we did not collaborate on a part of my work, but thank you a lot, you helped me always, especially to solve the technical problems I faced. Thank you all for being great mentors, and “Italian Language” teachers, I am sure I would not have learned Italian without you. Finally, I would say I had great memories with you that I will not forget in my life.

I would like to thank my friends Hoda, Rana, Mario, Ahmad, Mohammad, and Dr. Ali Ibrahim. We were not only able to support each other by deliberating over our problems and findings, but also happily by talking about things other than just our papers. In particular, I am grateful to Dr. Ali IBRAHIM for his support during the three years, he helped me in facilitating many things during my stay in Genoa. Apart from the lab and the research world, I am grateful for having new friends during my stay in Genoa; we shared unforgettable memories together and will share even much more in the upcoming years.

Last but not the least, my deepest gratitude goes to my precious family that tried to support and encourage me in every possible way. First, I am entirely grateful to the person who taught me the value of hard work in this life, to my father who passed away few months before I started my Ph.D. His absence did not stopped me; however, he gave me the persistence to continue what he started. He dedicated his entire life for his family, so I am dedicating this thesis for him even it represents nothing in return. Thank you so much “Dad”, I will never forget you, your soul is always guiding me even you are not present.

To my mother Fatima, my brother Abbass and my sister Mona, I am also grateful to you, for your tremendous moral and emotional support. Everything I have reached is because of your continuous guidance, encouragement and love. You are my everything in this life!

Table of Contents

1 Chapter One: Introduction

1.1 Context and motivation

Nowadays, with the growing usage of technology, computers and internet technologies become an essential part in all fields of our daily life. Computer systems are used in all fields, at schools, universities, hospitals, banks, governmental offices, etc. We cannot imagine the daily business operations around the world without the use of computer technologies. This technology controls a huge portion of the business process in any organization, and helps in speeding it up. The use of computers, storage, networking and other devices to create, store and exchange all forms of data has increased significantly in the last years. Interconnectivity and data generated by devices resulted in ‘an unprecedented improvement in the quality of life [1]. According to the Internet Domain Survey by ISC (Internet Systems Consortium) that attempts to discover every host on the Internet by doing a complete search of the allocated address space and following links to domain names, there are at least 1 billion host. The most recent survey available online is dated January 2019 [2].

Vinton Cerf, one of the "founding fathers of the internet", has announced that IPv4, or numerical addresses that allow the web to exist, are finishing. He said, “It is my entire fault, when we thought about the IP address system we thought of an experiment, and we believed that 4.3 billion addresses for an experiment would suffice”. Now, we just have to "migrate" to the new IPv6 protocol in a hurry, as are already doing ICANN (Internet Corporation for Assigned Names and Numbers) and various sites with significant traffic volumes, such as Facebook and Google.

Moreover, our daily life is becoming increasingly dependent on computers and networking technologies. Banking, shopping, and all manner of essential services are now available online. This deep dependence on the services available on computer networks underscores the need for providing security and reliable operation on these networks. Additionally, in recent years the number of attacks on computer networks has dramatically increased which consequently makes the security of computer networks even more important.

Cybercriminal activity is one of the biggest challenges that humanity is facing nowadays. According to the Official Annual Cybercrime Report by Cybersecurity Ventures, sponsored by Herjavec Group published in 2019, cybercrime is the greatest threat to every company in the world. Cybercrime will cost the world in excess of $6 trillion annually by 2021, up from $3 trillion in 2015; Cybersecurity Ventures predicted. For organizations, the costs associated with cybercrime are vast, they include damage of data, stolen money, lost productivity, theft of intellectual property, theft of personal and financial data, post-attack disruption to the normal

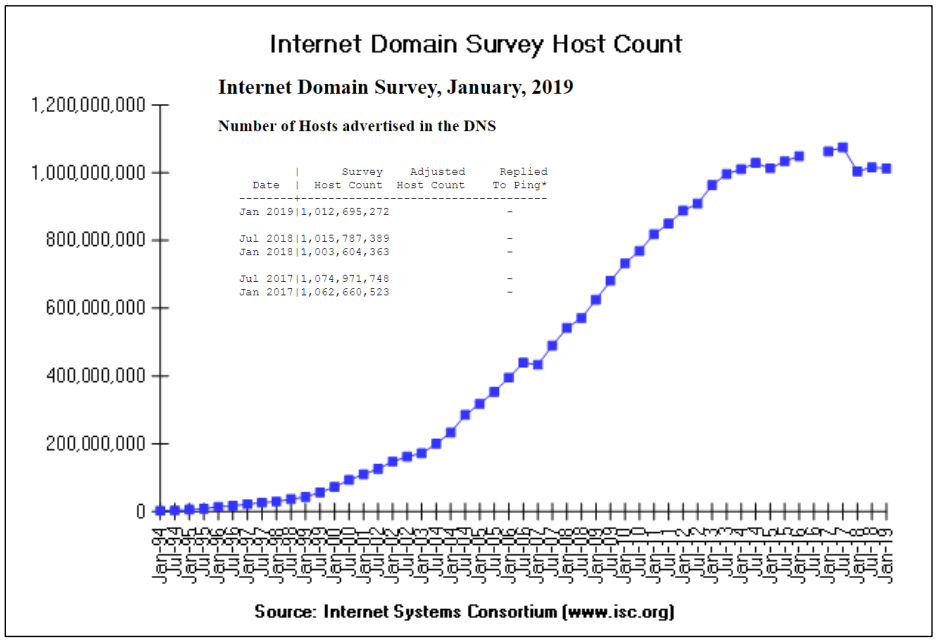

Figure 1. Number of hosts advertised in the DNS by July 2018

course of business, reputational harm, and much more [3].

Cyberattacks are crimes that are growing in a fast manner all over the world, and they are increasing in size, sophistication and cost. The Yahoo hack — which is considered one of the largest ever — affected 3 billion user accounts, and the Equifax breach in 2017 — with 145.5 million customers affected — exceeded the largest publicly disclosed hacks ever reported up until that time. These major hacks alongside the WannaCry and NotPetya cyberattacks, which occurred in 2017 are getting larger in scale and more complex than previous attack. Cybercrime is creating unprecedented damage to both private and public enterprises, and driving up IT security spending.

1.2 Problem Description

According to Cybersecurity Ventures that is one of the world's leading researcher for the global cyber economy, and a trusted source for cybersecurity facts, figures, and statistics. Cybersecurity Ventures predicts global spending on cybersecurity products and services will exceed $1 trillion cumulatively over the five-year period from 2017 to 2021. Taken as a whole, we anticipate 12-15 percent year-over-year cybersecurity market growth through 2021.

Worldwide spending on information security (a subset of the broader cybersecurity market) products and services will reach more than $114 billion (USD) in 2018, an increase of 12.4 percent from last year, according to the latest forecast from Gartner, Inc. In 2019, the market is forecast to grow 8.7 percent to $124 billion [4].

At the same time, the vast amount of data available about activities is giving rise to cybersecurity and privacy challenges. Data, such as technical and non-technical

documentation, financial and health records, and intellectual property may be worth millions of euros in the hands of hackers and data thieves. If organizations and companies do not address data security issues, critical threats to information privacy may develop. Businesses and other organizations thus must take action to secure the sensitive data they control [5]. With the diffusion of the internet and new storage media, data may be compromised on a larger scale and at a faster pace. With the sharing of data on networks, a threat to data security is becoming a major concern. Protection of information is necessary to establish and maintain trust between an institution and its stakeholders. Usually, data protection is treated as a technological issue to deal with; however, protecting data is not just a technology issue anymore [6]. Entire management systems inside companies now are giving enormous attention to organizational aspect. Policies, proved objectives, audits, training and awareness activities, compliance with legal and regulatory requirements for security and privacy have become important factors to be addressed in information security. One of the main requirements toward all of this stands the assessment of risk and its evaluation [7]. Consequently, organizations and in particular companies must realize the necessity of paying attention to the organizational aspects of data protection. Therefore, managing data protection can be better treated addressing two points of view: the organizational and the technological ones.

1.3 The PhD goal: Objective and contributions

This thesis describes and presents three years of work done at the "cybersecurity assurance and control" department at HITACHI Rail STS Company and at the University of Genoa. The work groups a set of solutions proposed, from the company point of view, to reach the main goal of the department, which is the data protection within the company. On the other hand, a research activity was also carried out at the university that aimed to classify web attacks using machine learning techniques.

The title of this thesis is "The importance to manage data protection in the right way: Problems and solutions", this title is chosen since the work represents a set of security solutions, which focuses on the management of data protection. It takes into account both the organizational and technological sides in proposing the approaches and methods, as stated in the subsection of the problem description. The work achieved can be divided into set of goals that are obtained particularly from the needs of the company and also the research community. These goals are organized into the following chapters:

The work done at the company:

- Chapter 3: A Data Protection Management Approach implemented by an innovative Information Security Management System.
- Chapter 4: The Information Security Management System.
- Chapter 5: A Comprehensive framework adopted in the security Risk management in Cyber Physical Systems.
- Chapter 6: Proposing guidelines to aid enterprises in selecting an applicable SIEM solution.
- Chapter 7: Risk Analysis Using PILAR software: An ISMS instructional document.

- Chapter 8: Conducting a Vulnerability Assessment Scan Using HPE Fortify Web Inspect: An ISMS instructional document

The work done at the University:

- Chapter 9: Classifying web attacks using network statistical features and machine learning: Analyzing the behavior of the attack generator.

The main part of the work was done at the Company; it handled the issue of managing data protection in a systematic way, through proposing a Data protection management approach (chapter three), and the shows the ISMS implementing this approach (chapter four) [8]. After that, a comprehensive framework is designed for the security risk management of Cyber physical systems, this framework represents the strategy used to manage the security risk management [9], and it falls inside the ISMS (chapter five).

In addition, inside the ISMS, another part of the work takes place (chapter six), suggesting the guidelines to select an applicable Security Incident and Event Management (SIEM) solution [10]. At the end, a summary for the professional work at the company is carried out, showing the way risk analysis (chapter seven) and vulnerability assessment (chapter eight) are applied, the work is demonstrated by instructional documents that will be part of the Company's ISMS.

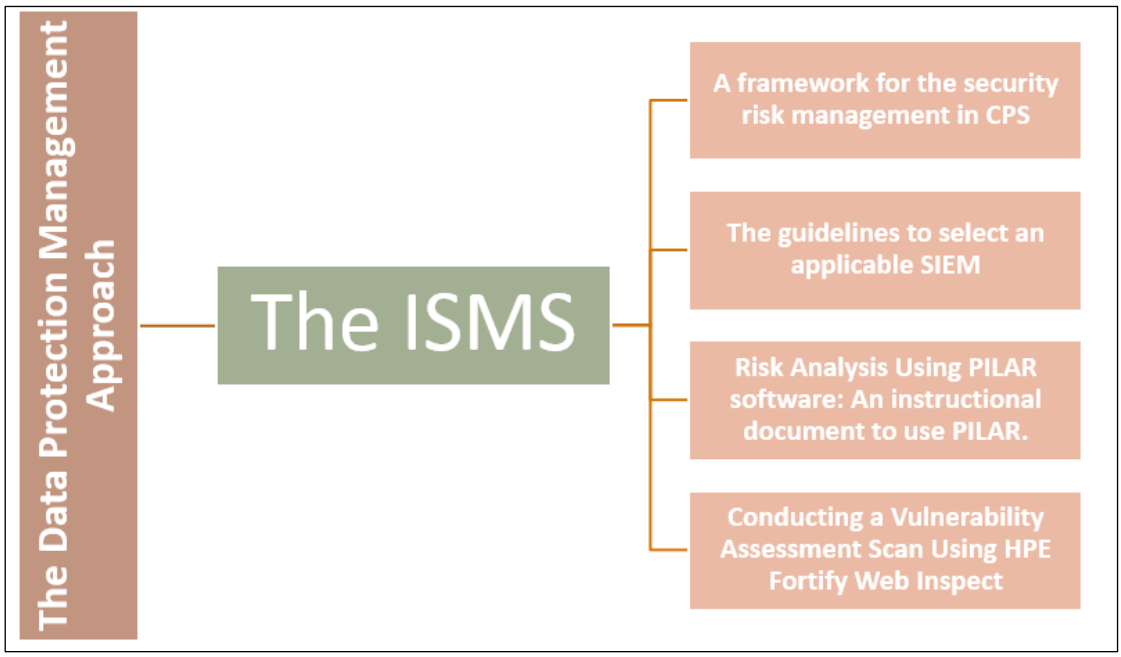

Figure 2. PhD Goals (related to the ISMS)

Figure 2 shows how the goals are related and falls all inside the ISMS implementing the data protection management approach which is the main objective of the CSAC department at the Company.

On the other hand, another research activity was accomplished to analyze the behavior of web-based attacks using machine-learning techniques. Chapter 9 explains this part of the work and shows all the steps done and results obtained.

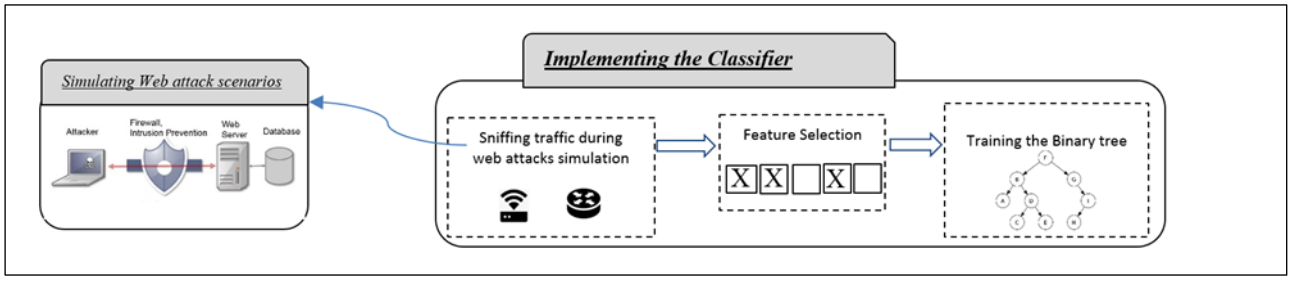

Figure 3. PhD Goals: A research activity of classify web attacks using machine-learning techniques

The goal is to get more useful information about the attackers that might help us in the mitigation and future prevention process, and to analyze the ability to detect the type of

software used in generating web attacks. The work starts by first creating a dataset of web attacks that addresses this side, labeled not only by the attack type, but also by the tool used to generate it. Then, a two-layers classification model is trained using this dataset to reach our goal mentioned above; where the first layer will be used to detect the type of the attack, while the second will prove the ability to classify the attacking tool, and though to be used in detecting the attacking tool type.

2 Chapter Two: Background and work aspects

Information and communication technology (ICT) has made remarkable impact on the society, especially on companies and organizations. The use of computers, databases, servers, and other technologies has made an evolution on the way of storing, processing, and transferring data. However, companies access and share their data on internet or intranet, thus there is a critical need to protect this data from destructive forces and from the unwanted actions of unauthorized users. To design a solution that truly protects the data, we must understand the security requirements relevant to our site, and the scope of current threats to our data [11] [12].

2.1 Hitachi Rail STS Company

This PhD thesis was done at the Department of “Cybersecurity Assurance and Control” at Hitachi Rail STS Company. Hitachi Rail STS is a leading company operating in the sector of high technology for railway and urban transport. The Company has the experience and resources to supply innovative transport systems for freight yards, regional and freight lines, underground and tramway lines, and standard and High-Speed railway lines. With an international geographical organization, the Company operates worldwide as lead contractor, system integrator and supplier "turnkey" of the most important projects of mass transportation in metro and urban railways. Hitachi Rail STS has a great experience in the design,

implementation and management of systems and services for signaling and supervision of railway and urban traffic.

2.1.1 The Cybersecurity Assurance and Control Department

- The main goal of this department is managing Data Protection of the Company from both organizational and technological sides, in terms of:
    - Defining policies, plans, procedures,
    - Performing risk assessment and treatment studies,
    - Vulnerability assessment and penetration testing,
    - Incident management,
    - Control activities as audits or approvals and assuring compliance with international standards.
- The department has its own strategy for managing data protection through an innovative Information Security Management System (ISMS) for managing Prevention, Monitoring, Detection and Reaction Phases.

To manage data protection correctly, it is important to take in account some aspects in the procedure of data protection, the main aspects are:

- Data classification levels.
- Threats and Vulnerabilities.
- Data security requirements.

2.2 Data classification levels:

Data is one of the strategical components of the corporate assets essential to a company. For this reason, it should be protected within a company in accordance with its own value and its significance to the company's business by implementing a classification process. Data

classification is also useful to identify who should have access to the technical data used to run the business versus those who are permitted to access test data and programs under development. Data classification must take into account legal/regulatory /internal requirements for maintaining confidentiality, integrity and availability. Data classification should define the following:

a. The owner of the information asset.

b. Who has access rights (need to know).

c. The level of access to be granted.

d. Who is responsible for determining the access rights and access levels.

e. Which approvals are needed for access.

f. The extent and depth of security controls.

However, data shall be classified by means of a method entailing an established structure of criticality and protection levels, which shall be determined in accordance with the potential impact on the company (e.g. the economic value, the damage to the company's reputation, the legal constraints and the strategical significance). An example of the classification levels defined in a decreasing order of criticality can be as follows:

- **CONFIDENTIAL:** Data concerning the company and/or its own subsidiaries, which may, when disclosed freely, cause an economical damage or affects the Company's - reputation.
- **RESTRICTED:** Data that can be freely accessed by the personnel working at the Company. This is the default classification level that shall be assigned every time a new piece of information is created.

- **PUBLIC:** Data that can be freely disclosed outside the Company, since its disclosure shall cause no damage to the Company itself.

In general, the data classification applies to confidentiality, integrity and availability. However, at Hitachi Rail STS, it was regulated only for confidentiality.

2.3 Threats and Vulnerabilities:

The two main kinds of security threats that affect a company are internal and external threats. Internal threats occur from within the organizations. This is probably one of the most dangerous situations because for instance co-workers may know passwords to access systems and are aware of how the systems are set up. Computers that are left unattended can be easily accessed by workers. And external threats are attacks done by hackers [11].

- *Internal Threats*: Previous research on cybersecurity has focused on protecting valuable resources from attacks by out-siders. However, statistics [13] [14] show that a large amount of security and privacy breaches are due to insiders. Protection from insider threats is challenging because insiders may have access to many sensitive resources and high-privileged system accounts. Similar style of exploitation is reported in [15] [16].
- *External threats*: External threats are those done by individuals from outside a company or organization, who seeks to break defenses and exploit vulnerabilities. Spying or eavesdropping, Denial of Service (DoS), Spoofing, Phishing, viruses, etc..., are all examples of external threats or cyber-attacks. However one of the emerging threats and the latest criminal invention is the Ransomware which is a special type of virus that does not destroy any data but simply encrypts all the data it finds on a PC with an encryption key that only the criminal has, and asks for money to give the key.

- *Vulnerabilities:* Weakness in the organization or company cyber-assets that a malicious attacker could use to cause damage. Vulnerabilities could exist in system, installed software, and network.

2.4 Data security requirements:

Data security concerns the use of a broad range of information security controls to protect the whole system (potentially including the data, the applications or stored functions, the database systems, the database servers and the associated network links) against compromises of their confidentiality, integrity and availability [11]. Data protection must address these main security requirements:

- ***Confidentiality***: This means that data must not be exposed to unauthorized individuals. In addition, access must be restricted to those authorized to view the data. Confidentiality has several different aspects: privacy of communications, securing storage of data, authentication of users, and access control.
- ***Integrity***: Data integrity means that data should be protected from corruption while it is stored in the database or transmitted over the network. Integrity has different aspects: only authorized users can change data, protecting the network and data against viruses designed to corrupt or delete.
- ***Availability***: Data must be available to authorized users, without delay. Denial-of-service attacks are attempts to block authorized users' ability to access and use the system when needed. Availability has different aspects: system resistance, performance, scalability should have adequate means [11].

Hitachi Rail STS adopts an information security strategy described in chapter 3 by a data protection management approach, and implemented by an Information Security Management

System (ISMS) which describes the organizational aspects of data protection inside the company, adopting the governance, risk and compliance approach, the ISMS is presented in chapter 4. From a technological point of view, Hitachi Rail STS adopts a defense-in-depth approach and maturity models to deploy the security controls in a prioritized and effective way in accordance to the organization's overall strategies and policies.

This thesis deals with more than one topic to reach the main goal, which is managing data protection. However, they all represents a part of the main cyber security strategy adopted by Hitachi Rail STS that is described by the data protection management approach (chapter 3) and implemented by the ISMS (chapter 4).

3 Chapter three: A data protection management approach developed by Hitachi Rail STS

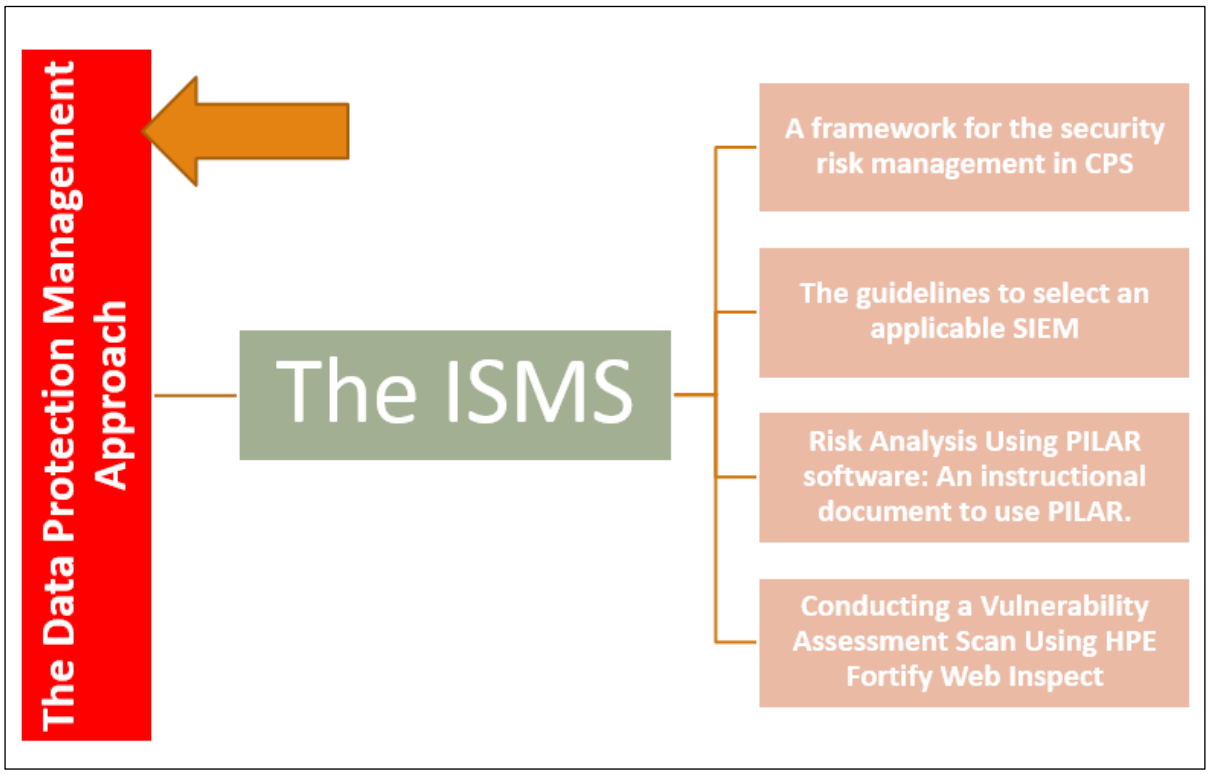

Figure 4. PhD goals: The Data Protection Management Approach implemented by an ISMS

Data has become the most important asset for an organization or company, and data protection is fundamental to make an organization achieve its success. Data loss due to a malware infection can lead to critical consequences. It is dangerous if companies' data fall into the wrong hands and it could be hard to recover, consequently data protection should be the key concern of companies and organizations. Most of the companies are connected to the internet for business reasons and this is potentially risky. Cyber-attacks, hacks and security breaches on the internet are no longer an exception [6]. The number of cyber-attacks is increasing, which may have a severe impact, economical too. Cybersecurity breaches can range from no or limited impact to Distributed Denial of Services (DDoS), stealing of data, manipulation of data, identity theft or even taking over control of systems and harm the physical world [17]. On the

other hand, some companies work on huge and critical projects that contain documentation that must be protected and not publicly disclosed. Data leakage or loss could lead to hazardous situations, so the confidentiality and integrity of data should be conserved. Companies must be prepared to protect the data: proactive plans and actions can reduce the effects of threats of any type. To reach this goal, it is better to adopt a good data protection management, which means having effective processes and methodologies in place to enable prevention, detection and reaction to any threat that could occur. Companies should give importance to actions, plans, polices, and address the organizational aspect and be aware and prepared to manage crisis situations. On the other side, they should work on technology, finding the best technological solution for each stage of the cybersecurity management.

In this chapter, we present solutions and key steps to manage data protection inside a Company from both organizational and technological sides. On the first side, by proposing an Information Security Management System that implements the cybersecurity strategy of the company through the three phases of cybersecurity: prevention, detection and reaction, and checks for compliance and improvement. And on the second side, by adopting a defense-in-depth approach and maturity models to deploy the controls in a prioritized and effective way against threat scenarios.

Data protection aims to protect the Confidentiality, Integrity and Availability (CIA) of company data and information whether it is processed, transmitted, stored on and/or in transit through networks and systems. Data protection involves the protection of all the cyber-space used in the company to store, process and transfer this data against unauthorized use, disclosure, transfer, modifications or destruction, whether accidental or intentional, or the loss of availability of these assets or business processes to authorized users. Data protection is

ensured by an information security (cybersecurity) strategy used to secure all assets involved in the storage, processing, transmission of data such as databases, computers, servers, network devices, etc.

The need for cybersecurity is becoming increasingly important due to our dependence on Information and Communication Technology (ICT) to store, process and transmit data. Companies do not want to be associated with cybersecurity hacks or viewed as having not taken appropriate security measures [17]. On the other hand, different types of threats and vulnerabilities that threatens company data varies between internal and external ones, and this requires different types of countermeasures that starts by setting plans, policies, complying to laws and standards, training internal staff, and so on, and also setting appropriate countermeasures. For this reason, data protection requires a strategy that covers both organizational and technological security aspects inside a company, applied on the cybersecurity phases of prevention, detection, and reaction (Fig 5).

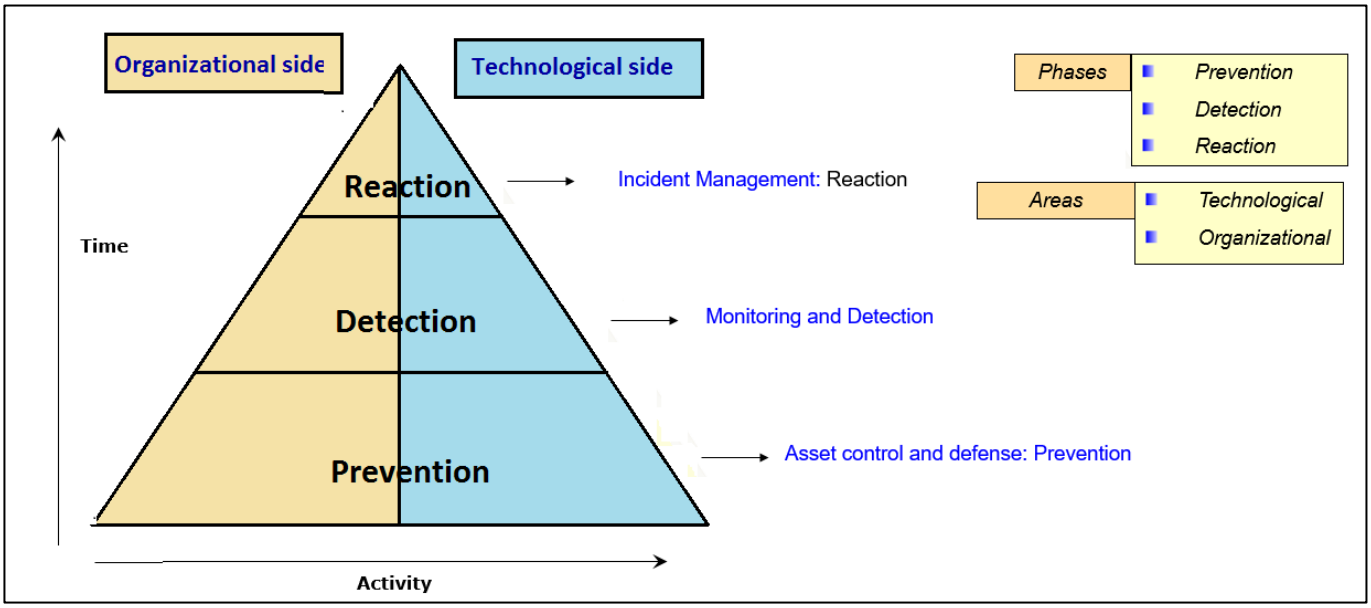

Figure 5. Data protection strategy process

From the organizational point of view, it is essential to set policies, actions, plans, responsibilities and ensure audits. On the technological side, the goal is to specify and implement the selected controls in the policies against threat scenarios. Both aspects should cover all phases of cybersecurity:

i. ***Prevention phase***: proactive phase for the defense of company assets.
ii. ***Detection phase***: monitoring of company assets.
iii. ***Reaction phase***: incident management.

In this chapter, we propose a data protection management approach aiming to protect the company's data both from the organizational and the technological side, which was inspired by the ISO 27001 requirements. This approach is implemented by an ISMS from an organizational aspect, and follows the Governance, Risk, and Compliance (GRC) framework. In addition, from the technological aspect, a defense in depth approach is adopted to deploy the controls selected by type and in a prioritized and effective way.

3.1 The organizational side: An Information Security Management System (ISMS):

An ISMS is a systematic approach for establishing, implementing, operating, monitoring, reviewing, maintaining and improving an organization's information security to achieve business objectives. The goal of ISMS is to minimize risk and ensure continuity by pro-actively limiting the impact of a security breach [18]. An Information Security Management System (ISMS) consists of the policies, procedures, guidelines, and associated resources and activities, collectively managed by an organization, in the pursuit of protecting its information assets.

The ISMS shall be balanced and integrated into the daily actions of employees; in addition, it shall be balanced among business goals, productivity and ensuring adequate data protection levels of the company and it shall ensure the privacy of employees. Business and IT staff which are relevant for information security activities shall be trained in order to ensure the application of the defined ISMS, and awareness initiatives shall be deployed to all employees. Next chapter describes in details the ISMS, its components and how it implements the information security process.

3.2 Technological point of view: Defense in depth approach

Data protection from a technological side is about executing the ISMS plans and operations, by selecting the right counter measures, specifying their types, prioritizing them by maturity levels, day-to-day operation, etc. To fully protect the data during its lifetime, each component of the information system must have its own protection mechanisms. The building up, layering on and overlapping of security measures is called defense in depth. Defense in Depth (DiD) is an efficient operational approach that enables to manage (with a risk-oriented approach) people, processes and technology. In IT environments, DiD is intended to increase the costs of an attack against the organization, by detecting attacks, allowing time to respond to such attacks, and providing layers of defense so that even successful attacks will not fully compromise an organization. A DiD strategy is necessary because of the new security threats and the importance of IT security monitoring of assets. Main variables that have increased the importance of DiD strategy definition are for example: the increased value of data, globalization, mobile working, virtualization, and decentralization of services. In this context, the company shall recognize the need to provide coordinated and multi-layered security architectures to mitigate security risks. Implementing DiD requires an understanding of

enterprise strategy, applicable internal and external threats, information asset classification, and technology supporting controls. The Defense in Depth strategy shall define all layers and technologies which, given company environment and security requirements, are necessary for all the different parts of the organization, as depicted in the figure below. Referring to Best Practice and Guidelines, Hitachi Rail STS adopted the DiD by using five levels to describe security actions based on the plans and policies of the ISMS.

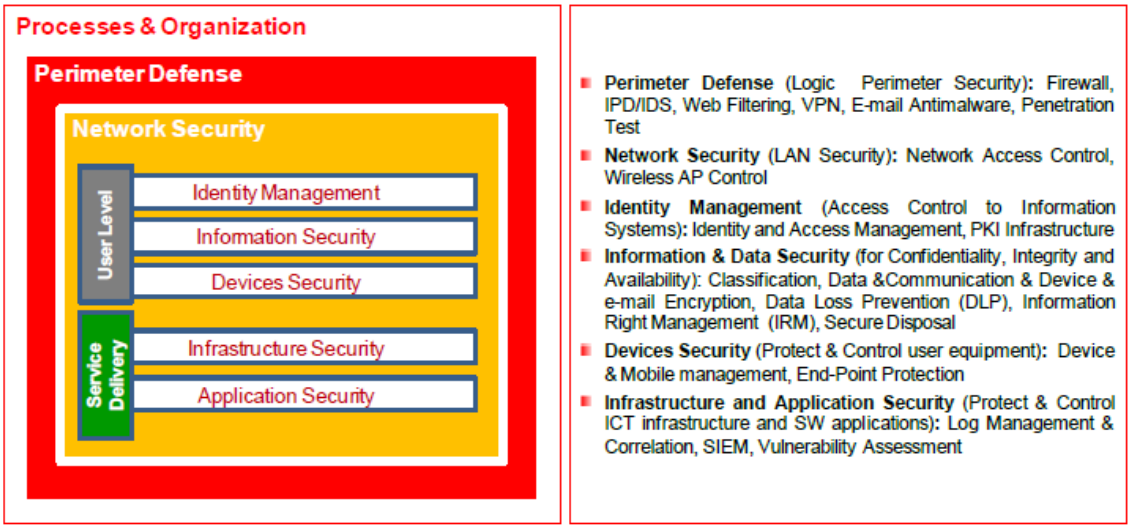

Figure 6. Defense in depth: Layering and setting technologies

Referring to Maturity Model, Hitachi Rail STS intended their policies to be applicable completely, although at a different pace, so they divided them according their complexity (to verify where controls would go deeper and where has been less deep). The maturity levels are used to set the level of security and control of a specific configuration or solution. The Maturity Model adopted by the company includes the following maturity levels:

L1: INITIAL

- Minimum set of acceptable security measures and control activities are designed and in place.

- Security measures and control activities have been documented and communicated to stakeholders and interested parties.

L2: IMPROVED

- Standardized controls with periodic testing for effective design and operation with reporting to management are in place.
- Improved or selected security measures are in place to harden specific controls or business areas.
- Automation and tools may be used in a limited way to support control activities and security measures effectiveness.

L3: OPTIMIZED

- An integrated control framework with real-time monitoring by management for continuous improvement (enterprise-wide risk management) of the security measures is in place.
- Automation and tools are widely used to support control activities and allow the organization to make rapid changes to the security measure in place.
- High level of security measures are available, addressing the trade-off between residual risk and costs.

4 Chapter four: The Information Security Management System (ISMS)

In this chapter, we describe the proposed ISMS, which is adopted by Hitachi Rail STS, and used to manage data protection inside the company. The ISMS implements the whole information security process used to protect the data within the company. Both information security and information technology departments are involved, and they have the responsibility and accountability of executing the sub process.

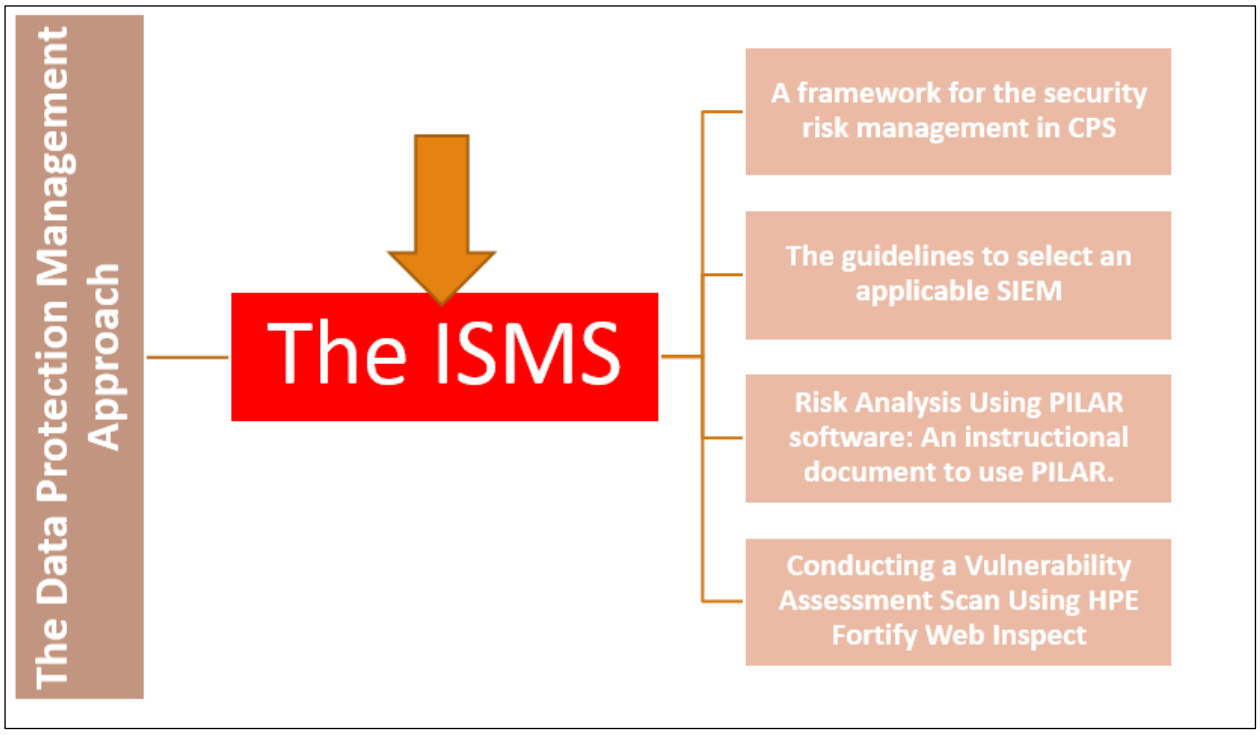

Figure 7. PhD goals (related to the ISMS): The Information security management system (ISMS)

Data must be treated as the core asset of any organization or company, which should be protected from all kinds of threats. However, to best protect the data, it is better to adopt a good management approach, which minimizes errors, saves time, increases awareness and prepares the company against incidents. In information security, taking due care of strategies, setting policies, plans, preparing and training staff, and complying with internal or external laws and standards should be given the same importance as operating technical tools. The focus should

be on both organizational and technical sides. Hitachi Rail STS Company gives a great significance to the organizational side of data protection, which is shown by its ISMS which was created in accordance with the international standards and frameworks.

Hitachi Rail STS implements an ISMS, and related documents are created, developed and published as means to implement data protection strategy, in accordance with the company business requirements, strategies, and relevant laws, regulations, and contractual agreements. The ISMS is based on a governance, risk and compliance (GRC) framework [8].

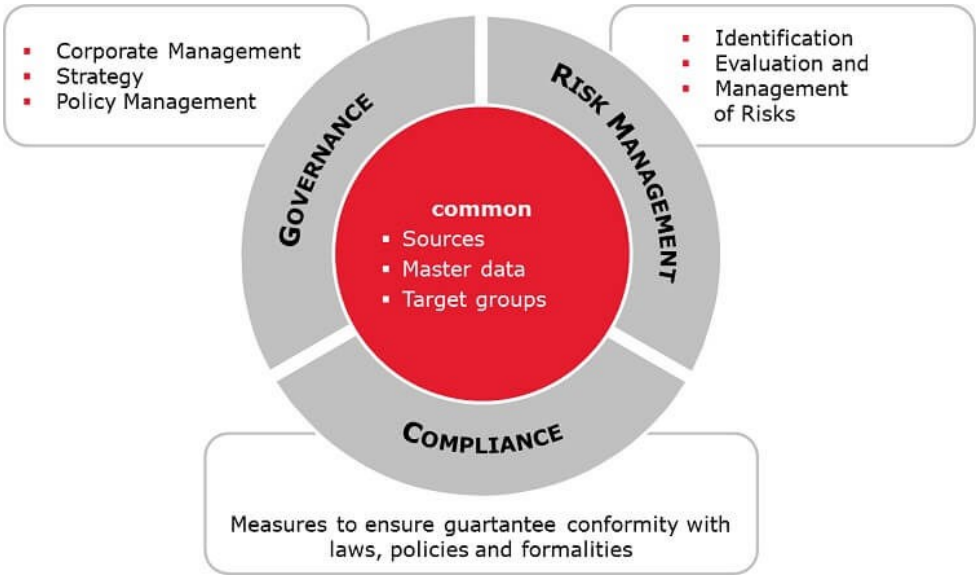

Figure 8. The GRC framework

a. **Governance:** Governance activities involve setting objectives to achieve and defining a way to achieve them while maintaining transparency with internal and external stakeholders. Governance tools, such as system controls and policies, are implemented in order to ensure that processes are followed in a proper manner.

   The Governance includes all activities necessary to define and implement a framework aimed at ensuring a proper data protection management. The main activities should define: roles and responsibilities, processes, policies and procedures (including supporting and

monitoring tools), audit plans. Main activities reported above must be executed according to the Segregation of Duties (SoD) principle. SoD aims at avoiding situations where a single person could execute or control several phases of the same process, or different processes identified as incompatible. The aim is to mitigate potential exposures to human mistakes or fraud events. Correct implementation of data protection strategy is verified through periodical audits performed by IS departments and by third parties.

b. **Risk:** Risk Management activities involve risk identification, assessment and mitigation plan definition. All risk management activities shall be performed on an ongoing basis in order to ensure that new risks are identified and previous identified risk are mitigated. Risk management acts as an internal control system that has to grow together with the business growth. The process of assessment, management and monitoring of risks through establishing and maintaining an appropriate risk framework is performed by the IS Manager. The assessment allows to identify the action plan to put forth in order to mitigate the identified risks and protect the Confidentiality, Integrity and Availability of assets.

c. **Compliance (Compliance to international standards and norms):** Compliance activities involve regulatory analysis in order to ensure the compliance with global and local applicable laws. Applicable laws are identified on the basis of the regulatory framework applicable to industry and country in which the company operates. Compliance can be oriented to internal policies and rules or to external laws and regulations, but in any case it represents a fundamental step in order to maintain the organization control inside its specific regulatory environment. In this context, compliance shall:

- Be maintained with all applicable national and international privacy legislation, and with international information security standards such as ISO/IEC 27001, GDPR (General Data Protection Manager) or other equivalent best practice/regulation required by the business.
- Ensure that all employees and third parties follow all security requirements.
- Ensure that all employees and collaborators as well as third parties with access to information systems are aware of their responsibility to report any security incident as quickly as possible.

4.1 The information security process implemented by the ISMS:

The activities of an information security (IS) process implemented by the ISMS can be divided into four vertical domains with different responsibilities and accountabilities represented by the company area in the figure below:

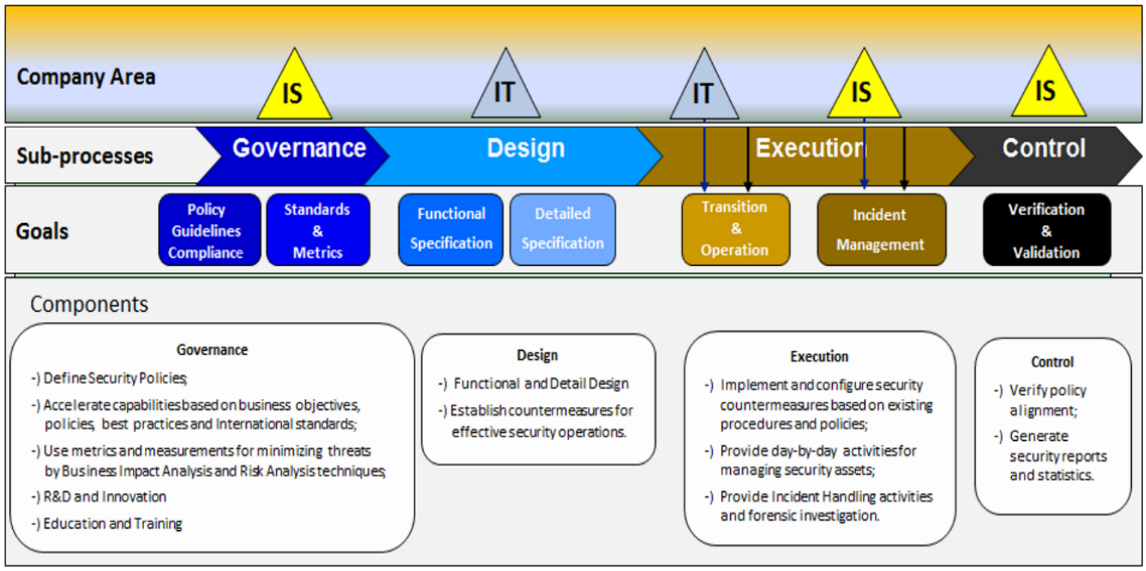

Figure 9. The information security process implemented by the ISMS for data protection management

Governance, design, execution and control are the four sub processes of the IS process implemented by the ISMS, where each sub-process has a set of goals:

-***Governance & Risk*:** defining strategy, policy (security levels), requirements (constraints), procedures, and conformity depending on internal or external requirements, laws and international standards. Then evaluate the Risk and identify the countermeasures to be taken to obtain an acceptable level of risk.

-***Design*:** defining information security architectures and technology solutions based on the countermeasures to be adopted and the approved budget in accordance with the defined strategy.

-***Operation (Operate and Execute)*:** putting the activities defined into operations, including the new transitions, the change of existing ones, day-by-day operations and maintenance of the equipment.

-***Control*:** assessing the adherence of current levels and security configurations to the policy, requirements and compliance set out in governance phase.

4.2 Segregation of duties

The ISMS is better developed in accordance to the "segregation of duties" principle as stated by the A6.1.2 control of Annex A in ISO 27001. The A6.1.2 control of the ISO 27001 states that conflicting tasks and areas of responsibility must be separated to reduce the chances of misuse, unauthorized or unintentional modification of the assets of the organization [19]. Each phase has a responsible department within the company, with the Information Security (IS) Department and the Information Technology (IT) Department being the involved departments.

- ***Governance***: The IS department has responsibility and accountability for this phase.
- ***Design***: The IT Dept. has responsibility and accountability for this phase.

- ***Operation***: The IT Department has the responsibility and the accountability for this phase. However, under this phase, the IS Dept. retains responsibility and accountability for the Incident Management task. In case an incident occurs, the IS department tries to understand the incident, find a solution and define the remediation, and finally gives the IT department the procedure to apply.
- ***Control***: The IS department has responsibility and accountability for this phase.

4.3 Documentation of the ISMS

At the end, the documentation of the ISMS should be also carried out. There are five main categories of documents used to describe the ISMS of Hitachi Rail STS Company.

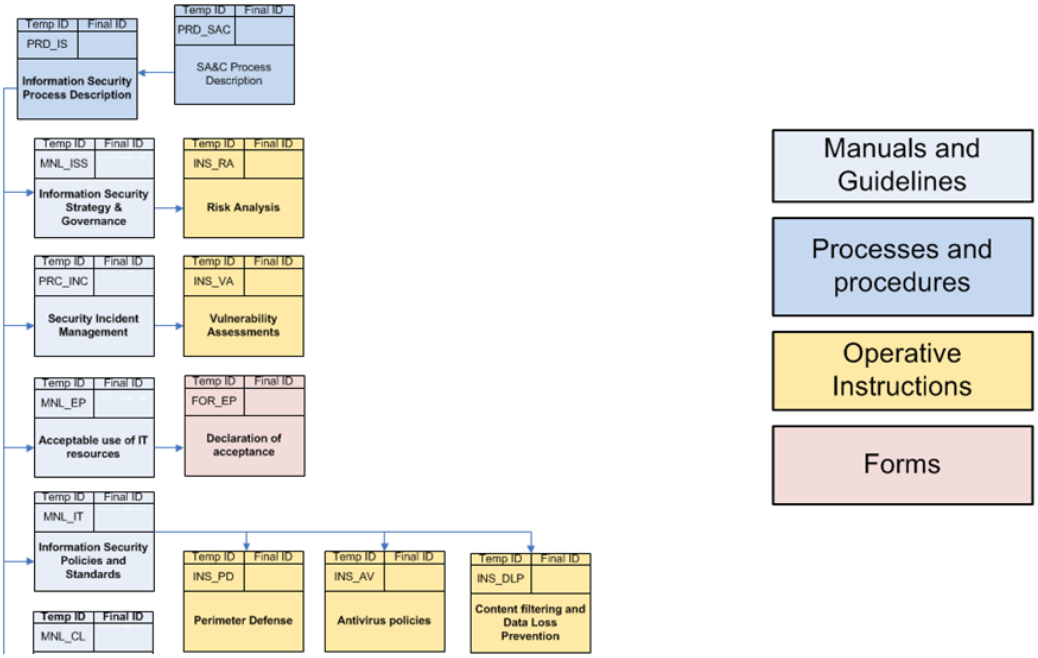

Figure 10. Documentation of different components the ISMS

Process Description (PRD): It is a high level describing the whole ISMS with – and includes all the responsible, accountable, consulted and informed departments involved in the ISMS.

Manual (MNL): This type of documents is a high-level directive. They describe the strategy, governance, and rules.

Procedures (PRC): These documents indicate who does what. They describe the responsible individuals and their duties.

Instruction (INS): These are the instructional documents. They give a brief description for the way of operation or installation indication how it is done.

Module - Template – Checklist (FOR): these are forms or check lists that must be filled for the purpose of requesting a service from the IT department for example.

5 Chapter five: A comprehensive framework to achieve a high common level strategy for the risk management of cyber-physical systems

There are several approaches followed to analyze and treat risks in the systems that rely on information technology. In general, these approaches try to measure or estimate the probability and the severity of the risks and at the end find solutions to minimize those risks. This chapter proposes an innovative framework designed to manage the risks in systems that depend on information technology. Inside the ISMS, falls this comprehensive framework. It is one of the strategies added to the ISMS and can an be followed in any risk management projects.

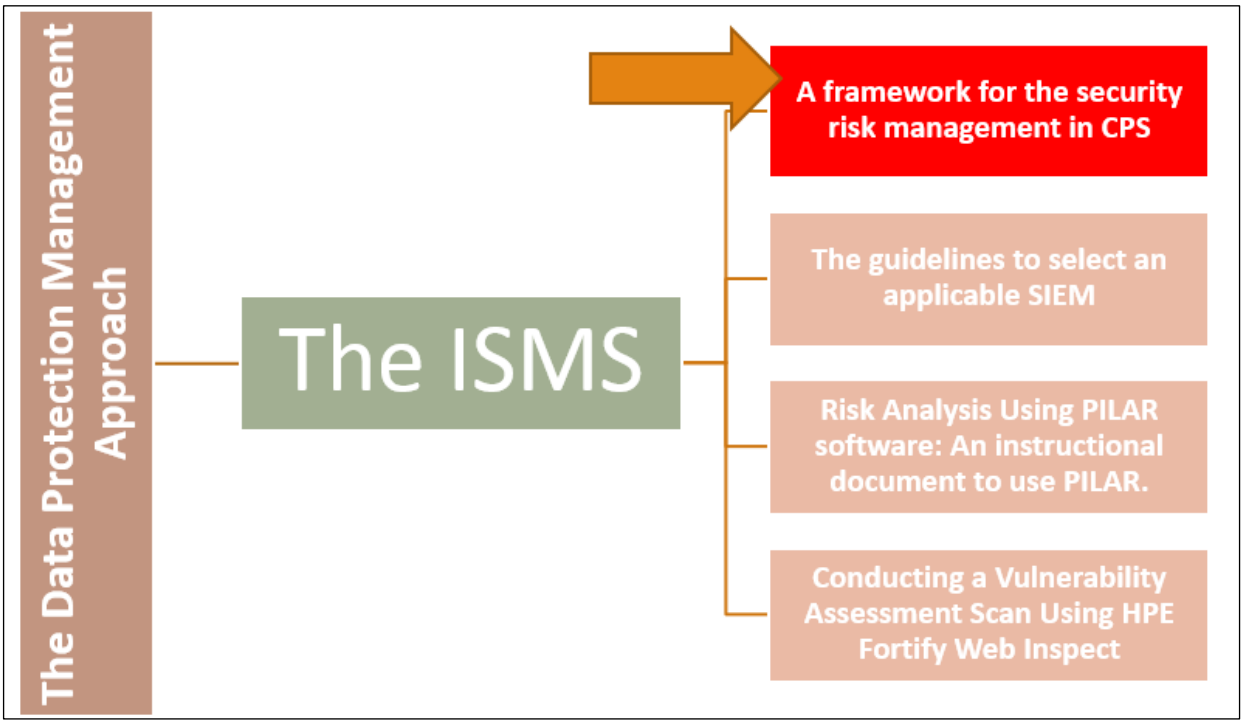

Figure 11. PhD goals (related to the ISMS): A framework for the security risk management in CPSs

Hitachi Rail STS Company works on projects in the field of railways, which are Cyber physical systems. A Cyber Physical System (CPS) is an intelligent complex system that combines diverse set of levels and types of computation and physical components, which are tightly integrated. CPS is facing huge security risks, especially cyber-attacks that can cause disruption

to physical services or create a national disaster. Information and communication technology (ICT) has made remarkable impact on the society, especially on companies and organizations. As a CPS relies basically on information and communication technology, this puts the system's assets under certain risks, and hence they must be kept under control by means of security countermeasures that generate confidence in the use of these assets. Therefore, there is a critical need to focus and give a great attention on the cybersecurity of these systems, which consequently leads to the safety of the physical world. An efficient risk assessment study applied to CPSs helps in the preparation against the crisis situations. This goal is achieved by adopting a solution that applies processes, plans and actions to prevent or reduce the effects of threats. Traditional IT risk assessment methods can do the job, however, and because of the characteristics of a CPS, it is more efficient to adopt a solution that is wider than a method, and address the type, functionalities and complexity of a CPS. This chapter describes a proposed framework that breaks the restriction to a traditional risk assessment method and encompasses wider set of procedures to achieve a high common level framework that could be adopted in the risk management process of cyber-physical systems.

A cyber-physical system refers to the systems that combine both cyber and physical resources, where there is a strong relation and coordination between these resources. Such systems are controlled or monitored by computer-based algorithms, tightly integrated with the internet and its users. CPS is basically a control system with distributed networked, adapted and predictable, real-time, intelligent characteristics, where human-computer interaction may exist. It is widely used in critical national infrastructure, such as electric power, petroleum and chemical and so on [20]. Moreover, many urban transportation and railway systems around the world have deployed some form of communications-based automatic train control (e.g., [21], [22]). In

those systems, multiple cyber components, including wireless communication. The potential implications of this evolution could be multi-faceted and profound, especially when it comes to the issue of security. If such systems were subject to a physical or cyber threat, the consequences will be unimaginable. These systems are susceptible to different types of risks related to information systems vulnerabilities. No one doubt about the hazardous consequences that would occur in case a malicious software succeeds in controlling the system, i.e. any fail in systems controlling drive-less metros will lead to huge loss. Security breaches in the cyber domain, such as falsified information or malicious control logic, can have a complicated impact on the physical domain [23]. "The cyber breach will lead to complicated physical consequences". Cybersecurity breaches can range from no or limited impact to Distributed Denial of Services (DDoS), stealing of data, or even taking over control of systems and harm the physical world [17]. In power industry, the monitoring system of a power plant was attacked by unknown sources in 2010, which leaded to 900MW load loss in 7 seconds. In energy industry, the computer system of Iran Bushehr nuclear power plant was invaded by "Stuxnet" in 2010, leading a serious chaos in the automated operation of the nuclear facilities and a serious setback of Iran's nuclear program. In transport service, in the network for managing and monitoring the operation of the Shinkansen, due to an exception in the management system of control schedule, signaling and line switching point in 2011, Japan's 5 Shinkansen operation management system encountered failure, 15 trains were in outage, 124 trains were delayed and 8.12 million people's travel were affected. In this way, we can conclude that CPS security is so important that risk incidents in the system may affect national security and stability. Taking all these security incidents seriously, we conclude that any attack

in the cyber layer of the cyber physical system could lead to hazardous situations and even to loss in lives [20].

There are several approaches for the problem of risk assessment and treatment: informal handbooks, methodical approaches or supporting tools, where all provide a guide for risk assessment and treatment. However, methods might differ in some steps, or in the way of identifying and valuating the assets or threats. Some are basically used in cyber security of information systems, and others can be used in physical security. Many of the proposed solutions tries to measure or estimate the probability and the severity of the risks after identifying the assets and threats using traditional IT risk assessment methods. Some of these solutions did not address the characteristics and the complexity of CPS, which needs a broad range of management. The great challenge of these approaches is the complexity of the problem they have to face; complexity in the sense that there are many elements to be considered and, if it is not done rigorously, the conclusions will be unreliable.

Hitachi Rail STS is a leading Company operating in the sector of high technology for Railway and Urban Transport. Hitachi Rail STS has a great experience in the design, implementation and management of systems and services for signaling and supervision of railway and urban traffic [24].

Hitachi Rail STS believes that there is a critical need to adopt a comprehensive strategy for the problem of applying risk management study to a cyber physical system. However, the complexity of the CPS is greater and such systems need more procedures to be performed. So this chapter proposes a framework that aims to reach a common high level solution, it is different and broader than a traditional IT risk management methods where their goal is mainly focused on identifying and measuring the severity of the risks and try to reduce it to an acceptable extent. However, it encompasses seven steps and inspired by the PDCA cycle, and

centered upon the cyber side and its assets; however, this doesn't mean that the physical assets are out of the frame, as the physical assets of a CPS are mostly controlled by others in the cyber side. This framework is characterized by a set of procedures that starts by modeling the system's assets and functionalities, selection of potential threats to the CPS, conducting risk assessment and treatment through a methodical way, vulnerability assessment, ensuring the compliance with global and local applicable laws, and finally applying maintenance and improvement activities. This chapter presents a set of aspects that the approach mentions, and then describes the proposed framework and a case study that shows how Hitachi Rail STS Company applies this framework.

5.1 Aspects and Requirements:

5.1.1 Cyber physical system security:

CPS security has some distinct characteristics as it is different from traditional IT system. In traditional IT system the first important aspect of information security is confidentiality. Confidentiality means the protection of data, providing access for those who are allowed to see it while disallowing others from learning anything about its content. However for CPS, the availability comes first, then the integrity and confidentiality.

CPS has more attack points and fault points than IT system. Any safeguard measures shall not interrupt the response to the physical system or delay the response. In traditional IT system access control can be deployed without affecting the services of IT system. In CPS all these measures should be discussed and tested to great details. The data flow shall not be hindered or interfered. CPS is a system of systems, the tight coupling between the physical system and cyber system has led to potential cascade effect of the whole system. Malfunction whether in

cyber part or in the physical part will spread to other part of system. The attack points of CPS are no longer confined to cyber-attacks from the Internet [20].

5.1.2 Dependencies and Accumulated risk:

As mentioned above, it is more efficient for a security strategy to start with functional modeling of assets, by defining relations and dependencies, as it leads to more precise and coherent study. Dependencies affect all the calculations done to assess the risk. Since assets depend on each other, the occurrence of threats on assets causes a direct harm on them and an indirect harm on others that depend on them.

5.1.3 Security requirements:

The cyber security of CPS corresponds the use of a wide set of security controls to protect the whole system against compromises of their confidentiality, integrity and availability. The cybersecurity of CPS must address these main security requirements:

- Integrity: It means that only the authorized users can change in the assets, it is satisfied if the assets are not changed by an unauthorized party.
- Confidentiality: This means that the assets must not be exposed to unauthorized individuals. And access must be restricted to those authorized. This is satisfied if the assets are not read or accessed by an unauthorized party.
- Availability: is satisfied if the assets or services are available and without delay.

If the system was exposed to malicious activities, consequently physical components will be affected and even damaged. It can be said that in a CPS, the availability comes first, then the integrity and confidentiality. It is quite simple that if the cyber system was under attack, the physical processes might be no longer under control and the consequence will be disastrous.

5.2 Related Work

Traditional IT risk assessment methods can achieve part of the work to mitigate the risks in a CPS, however, and because of the characteristics of a CPS, it is more efficient to adopt a solution that is wider than a method, and address the type, functionalities and complexity of a CPS. Some of those methods include MEHARI, EBIOS and MAGARIT.

5.2.1 MEHARI

MEHARI (MEthod for Harmonized Analysis of RIsk) is an Risk Assessment (RA) and Risk Management (RM) method, it is an Open Source and free methodology integrated and achieved for the assessment and management of risks associated to information and its treatments. MEHARI is developed and updated since 1996 by CLUSIF and CLUSIQ. It includes, directly in its knowledge bases, many formulas for the direct assessment of risk and selection of the ways to reduce them. MEHARI gives indications for the business stakes identification and valuation, the resulting classification of assets (according to ISO 27005) for the Availability, Integrity and Confidentiality security criteria is performed. Also the likelihood of the various threats is identified and the evaluation of the security measures to reduce the risk may be collected from audit questionnaires. All the elements for risk evaluation are available for the next phases, MEHARI knowledge bases provide comprehensive lists of risk scenarios associated with the assets and the various threats. The combination of stakes, threats and vulnerabilities included in the method allows analyzing the risks situations and preparing for the risk assessment.

For risk assessment, using MEHARI the critical risks may be displayed and analyzed under various other forms: by asset criterion, type of threat, actor, etc. For risk treatment, the risk managers or auditors have the capacity to select the treatment option (reduce, accept, transfer/share, avoid) and, in the case of a decision of reduction,

MEHARI provides the capability to select the additional security measures for the reduction of likelihood and/or impact and to integrate them in additional projects depending on the level of the resources and types of organization [25].

5.2.2 EBIOS

EBIOS (**E**xpression des **B**esoins et **I**dentification des **O**bjectifs de **S**écurité - Expression of Needs and Identification of Security Objectives) is a method for analysis, evaluation and action on risks relating to information systems. It generates a security policy adapted to the needs of an organization. The method was created in 1995 and is now maintained by the ANSSI, a department of the French Prime Minister.

EBIOS is a comprehensive set of guides (plus a free open source software tool) dedicated to Information System risk managers. Originally developed by the French government, it is now supported by a club of experts of diverse origin. This club is a forum on Risk Management, active in maintaining EBIOS guides. It produces best practices as well as application documents targeted to end-users in various contexts. EBIOS is widely used in the public as well as in the private sector, both in France and abroad. It is compliant with major IT security standards.
EBIOS gives risk managers a consistent and high-level approach to risks. It helps them acquire a global and coherent vision, useful for support decision-making by top managers on global projects (business continuity plan, security master plan, security policy), as well as on more specific systems (electronic messaging, nomadic networks or web sites for instance).
EBIOS approach consists of a cycle of 5 phases:

- Phase 1 deals with context analysis in terms of global business process dependency on the information system.

- Both the security needs analysis and threat analysis are conducted in phases 2 and 3 in a strong dichotomy, yielding an objective vision of their conflicting nature.
- In phases 4 and 5, this conflict, once arbitrated through a traceable reasoning, yields an objective diagnostic on risks. The necessary and sufficient security objectives (and further security requirements) are then stated, proof of coverage is furnished, and residual risks made explicit [26].

5.2.3 MAGERIT

See chapter 7.

However, the common point between these methods is that they do the job in the traditional way and might not be suitable for systems like CPS. They start by identifying risks, analyzing and evaluating them, then trying to mitigate by risk treatment phase, then acceptance and communication stages. And as mentioned before, those traditional risk assessment methods lacks for some essential procedures such as system functional modeling or asset modeling, threat modeling or selection, vulnerability assessment, compliance, an implementing-safeguard approach, and at the end the maintenance and improvements, which are all included in the proposed framework in this chapter. These procedures included in the proposed framework, might aid in the risk management of systems like CPS, which are considered complex and need to be managed in a very structured manner, and assert successful performance.

5.3 The proposed comprehensive framework for the risk management in CPS

Commonly, when there is a need to assess risks, traditional methods are used to do the job. Traditional risk management methods involve the following steps: risk identification, assessment and mitigation plan definition. However, a well-designed risk assessment of CPS will provide an overall view of CPS security status and support efficient allocations of safeguard resources. Though traditional IT system risk assessment is quite mature, a distinct

risk assessment method for CPS is needed to cover the growing security issues due to the large differences between IT system and CPS [20]. This chapter presents a framework inspired by the PDCA (PLAN-DO-CHECK-ACT) cycle. The proposed framework adds a boarder set of procedures for a traditional risk assessment method.

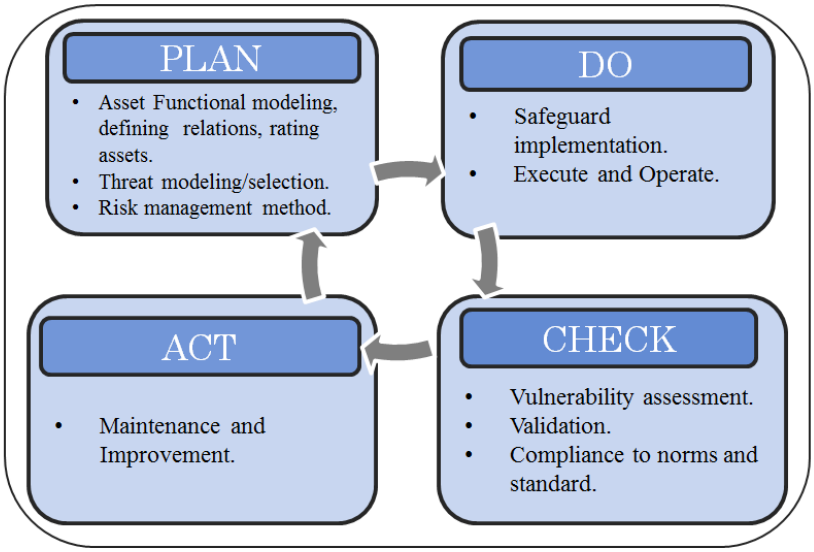

Figure 12. The proposed framework inspired by the PDCA cycle

In order to assure compliance with Security and safety requirements, Hitachi Rail STS needs to define and adopt a holistic framework for Risk Assessment and Treatment activities of CPSs. Figure 12 show how each step of the framework falls inside one of the phases of the PDCA cycle. It is a divided into the following seven steps:

i. System Functional Modeling
ii. Threat Selection and Modeling
iii. Applying Risk Management method (Assessment and Treatment plan)
iv. Safeguard implementation: Appropriate Security Governance model
v. Vulnerability assessment
vi. Compliance and Validation
vii. Maintenance and Improvement

To ensure the continuous improvement, the framework is based on Deming PDCA Cycle where each phase, because of the complexity of a CPS, can by divided further in a few steps (Fig.13 shows the 7 steps). The steps are applied in order: starting with the "PLAN" phase, first step is "***System Functional Modeling***" which designs the model for the CPS system showing the functionalities, dependencies, relations between the assets and defines also rules and Acceptable Risk Levels. Then the second step is "***Threat Modeling and Selection",*** it selects the potential "threats" that match the CPS's assets: this can be done by referring to historical data such as reports, statistics, observations, logs, etc. In particular to execute these actions a dedicated commercial tool, called RMAT, has been identified and adopted. Finally, always in PLAN phase, the first two steps are the input to the "***Risk Management***" step, where an appropriate method is selected to assess the risk (Risk Assessment) and helps in selecting the appropriate measures for keeping the risks under control (Risk Treatment). For performing this job, Hitachi Rail STS has identified and adopted a commercial tool, named PILAR, that implements a method called MAGERIT which is suggested by the European Union Agency for Network and Information Security (ENISA).

After that, "S***afeguard Implementation***" takes place, reflecting the "DO" phase of a PDCA, where the chosen decisions in the Plan phase are put into operation. At Hitachi Rail STS, the Defense in depth (DiD) approach is adopted while implementing safeguards: this approach, based on layering, helps in faster detection and slowing down attacks.

Afterwards there is the CHECK phase, represented by the "V***ulnerability Assessment and Penetration Test"*** process: it plays a key role in revealing the vulnerabilities yet present on the system and not protected by already installed safeguards. Because a CPS contains various set of HW/SW assets such as network appliances, servers, end-points, applications, web services, databases, etc., the Vulnerability Assessment and Penetration Test activity is applied basically

on 3 levels: Application, Network and Operation System Levels. Based on all previous findings and evidences, the CHECK phase is completed by a compliance control to ensure complying of the system to security best practices or international standards, e.g. ISO/IEC 27001/27002. Finally, the Deming Cycle is concluded by the ACT phase which contains ***"Maintenance and Improvement"*** activities to correct and improve the system.

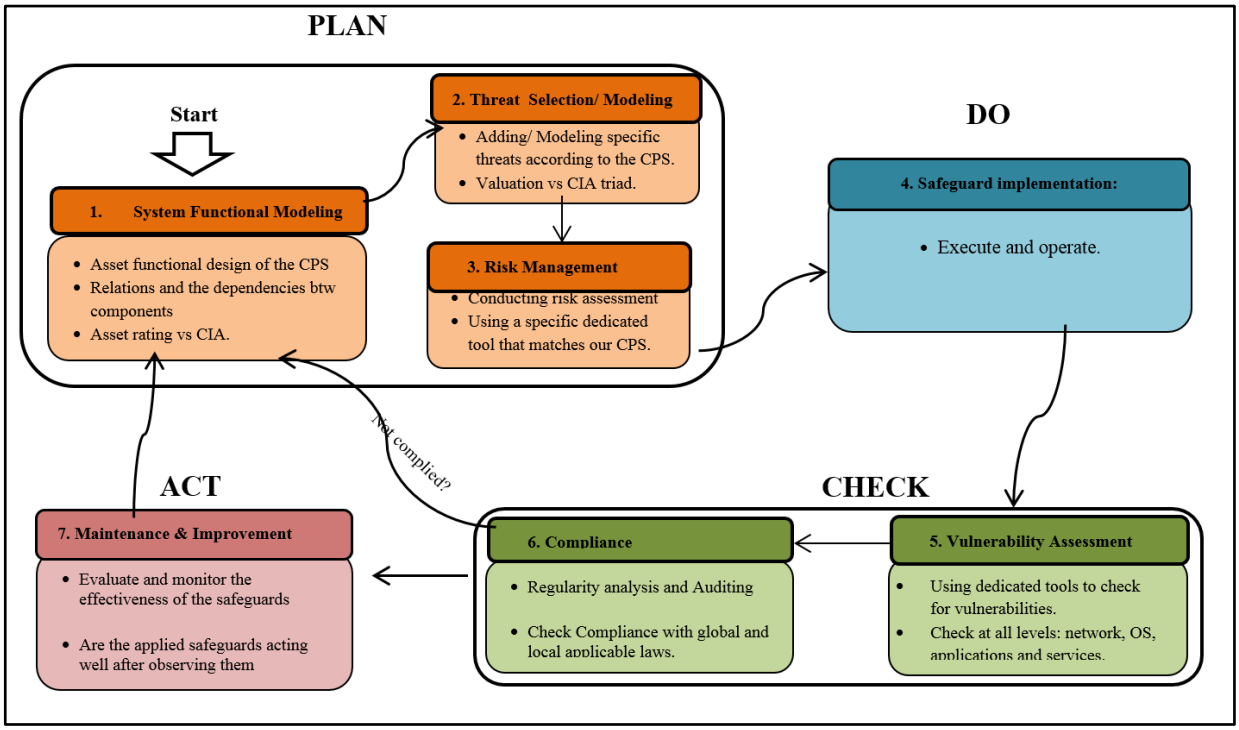

Figure 13. Hitachi Rail STS's framework for the Risk Assessment and Treatment of Cyber Physical system.

5.3.1 System functional Modeling (Asset Modeling):

Creating a system functional model has a great impact in showing the structure and the components of the CPS, and in demonstrating the relations and the dependencies between the different assets, and so to have a clear and precise simulation for the system in real life. It is the step where the whole framework depends on, in this stage it is meant to model the physical and cyber components and their interactions and operational characteristics. Asset Modeling can be considered as the most important step in this approach. This step must be done first with the owners of the system. The scope of this part is to help the system owners or information

sources in creating a system functional model and in the valuation of the system's assets. For this task, two steps are followed:

a) Creating a functional model for the system, which is a structured representation of the system's components (assets) and functions (activities, processes, operations).
b) Rating of the assets (based on CIA) using criticality levels and according to the consequences on CIA that would happen case of their protection failure.

The two steps must be done by the owners or under the supervision of them. In this way, a typical representation or a general view for the system is carried out which aids in the risk management study.

5.3.2 Threat Selection/ and Modeling:

Each CPS differs by the services and functionalities that it offers. Threats vary from one system to another, based on the available assets and their level of valuation. Different cyber physical systems means different assets and though different types of threats. Threats can be grouped and associated to homogenous group of assets called asset classes. Threat selection is about understanding the most suitable threats that are expected to happen and matching them with the different asset classes of the cyber physical system. The appropriate threats-to-assets should be selected in this step to be fed into the "Risk Management study" step, and should be applicable to the assets presented in the previous step. Mainly cyber-security threats are covered; that is, threats applying to information and communication technology assets, but additional non-IT threats could also be included in order to cover threats to physical assets that

are necessary for the operation of the CPS. This work can be done by referring to historical data such as reports, statistics, observations, logs, etc.

Figure 14. Common threats for the "Threat selection and Modeling" step in CPS

The ENISA Threat Landscape provides an overview of threats, together with current and emerging trends. It is based on publicly available data and provides an independent view on observed threats, threat agents and threat trends. Over 140 recent reports from security industry, networks of excellence, standardization bodies and other independent institutes have been analyzed [27] ; the figure above shows a sample for threats that threaten cyber physical systems. However, risk analysts are responsible for selecting and valuating the appropriate and expected threats that are likely to occur and match the system's assets. First, the general model is obtained by experts, reports, statistics, and then threats that match the context, type of the CPS and the given assets are kept and fed to the next step. Threat Modeling eases the risk analysis study in various ways; mainly it prepares a wealthy and substantial threats-to-assets convenient dataset that fits a case study. There are dedicated tools that help in threat modeling, and section 5.3.2 shows one of them which is used by Hitachi Rail STS Company.

5.3.3 Risk management plan:

Risk management is divided into risk analysis and treatment, risk analysis is a systematic process for estimating the risks to which the system's assets are exposed. Risk management is the part of planning, while in treatment is where decisions are taken. These decisions are demonstrated and established in the implementation step.

a. Risk analysis: A risk can be defined as the degree of exposure to a threat that may occur on one or more assets causing damage, it is an indicator of what could happen to the assets if not properly protected. It is important to know what features are of interest in each asset and to what extent these features are in danger, that is, analyze the system. There are several methods and ways for the problem of analyzing the risks: informal handbooks, methodical approaches or supporting tools, where all provide a guide for risk analysis. However, methods might differ in some steps, or in the way of identifying and valuating the assets or threats. Some are basically used in cyber security of information systems, and others can be used in physical security. Risk analysis study must be applied using an appropriate method and tool for the risk analysis step in the cybersecurity of CPSs. Applying a risk analysis study includes:

   i. Identifying and classifying assets by types, establishing dependencies between them and evaluating them according to security requirements
   ii. Identifying and valuating threats and their likelihood.
   iii. Identifying current safeguards and valuating them according to the level of effectiveness.
   iv. Evaluating the risk on the CPS where valuations for assets, dependencies, and threats are all involved in the calculation.

b. **Treatment plan:** On the other hand, this sub-step must also carry out the risk treatment activities that should be applied. Risk treatment activities allow a security plan to be prepared which, when implemented and operated, meets the proposed objectives with the level of risk accepted by the Management. In the treatment plan, the right counter measures are selected with types, and then prioritized. Moreover defining their cost/complexity, effectiveness and efficiency metrics must be also addressed. The objective is to deploy the controls selected by type and in a prioritized and effective way. For example, same safeguard can contrast more threats at the same time and overlapping/redundant safeguards must be avoided. However, sometimes, when a series of safeguards are in place and the management process is mature to a certain extent, the system will still be exposed to a risk called "residual".

5.3.4 Safeguard implementation: Operations

This step deals with the implementation of security plans and decisions taken in the treatment plan, it takes as input the activities defined and puts them into operation. It also deals with the technical side, and defines the best technological solutions based on the countermeasures to be adopted and the approved budget in accordance with the defined strategy. Implementation of safeguards must ensure the availability and the capability of the organizational staff to manage the tasks scheduled to implement them, as well as other factors, such as the budget of the organization, relations with other bodies, legal, regulatory or contractual changes, etc.

Even when the risks have been treated, residual risks will generally remain. Residual risk means that the current level of risk is accepted and is under a "careful chosen" threshold, where trying to eliminate it could be extremely expensive. So applying security patches and ensuring

the secure configuration of all appliances is maintained continuously, also assets are monitored and logs are analyzed to detect any improper actions.

5.3.5 Vulnerability Assessment:

Vulnerability is a weakness in the assets that a malicious attacker could use to cause damage. Increasingly sophisticated tools help to penetrate existing network connections. After implementing the safeguards in the previous steps, a vulnerability management process is needed to check if the assets of the cyber physical system are really still exploitable to threats. At the technical level, this is done by vulnerability exposure tools, with simulation of attack paths (similar to MITRE attack matrix). The end result can be patch management or better, in some complex environment, virtual patching (i.e. putting layer of defense that stop the attack before it reaches the endpoint, without the need to change configs of the endpoint itself). Furthermore, log analysis could be useful in revealing vulnerabilities; it requires much expert, knowledge, and maybe time consuming to do manual log analysis. At the end, when detecting irregular issues, it is required to return to the iteration cycles for proposals and solutions.

5.3.6 Compliance:

Assessing the adherence of security configurations to the policies, requirements and regulations are set out in this stage. Compliance activities also involve regulatory analysis in order to ensure the compliance with global and local applicable laws based on the requirements, or even with respect to verification schemes to be achieved or maintained. And in case of non-compliance, it is required to return to the iteration cycles for proposals and solutions.

5.3.7 Maintenance and Improvements:

Finally, the evaluation of the effectiveness and efficiency of the applied safeguards is measured to achieve the needed improvement and maintenance. It is recommended to deploy some

elements that allow controlling the measures implemented in order to assess their effectiveness and to have an insight about them to figure out if there are new problems or there is a need to update their level.

5.4 Applying the security Risk Management proposed framework for a CPS

This section shows how the proposed framework is applied at Hitachi Rail STS Company. Each subsection describes the procedure followed. The 7 steps are demonstrated below, showing how they were applied to achieve this overall high level framework of Risk analysis and treatment for CPS.

5.4.1 System Functional Model

The first step is to design a functional model for the system, i.e. it is fundamental to define the scope of the system, the basic components forming the CPS and their composing assets (physical and cyber), and also establishing the relations and dependencies between them. This step is done based on information coming from the owners, since they are familiar and have the knowledge about their system. The functional model will be used to rate the assets against the basic security dimensions Confidentiality, Integrity and Availability (CIA traid), as shown in the figure below:

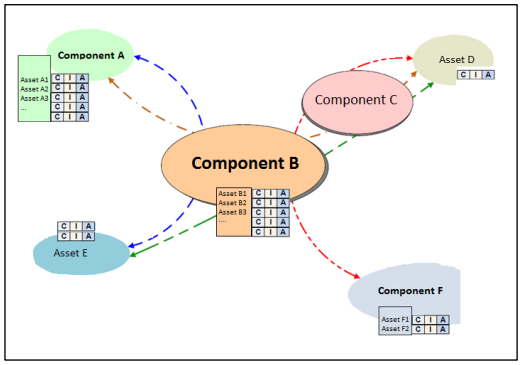

Figure 15. A system functional model example for the CPS

Then provide a high level asset rating for each with the assistance of the system's owners and based on the tables defined below. Figure 16 gives an example of the asset's security

dimensions rating, where each asset has a triad rating that represents respectively the confidentiality, integrity and availability rate.

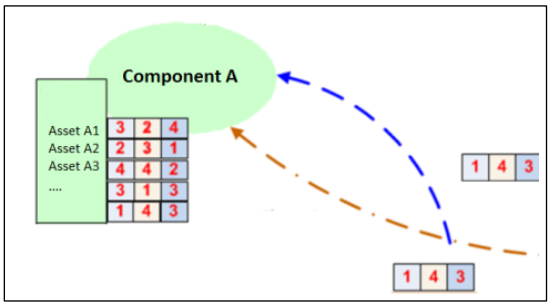

Figure 16. Rating assets

The assets' rating is carried out on each security dimension. Rating represents a pre-valuation step for the assets, where criticality levels are defined. These levels will be used with a scale from 1 to 4, where "1" describes the lowest critical level and "4" is the highest. And so, each security dimension gets one of the four levels representing the rate value. For each level, a description is given that helps in choosing the suitable asset's level. The three tables below explain the levels of rating according each security dimension.

Table 1. Asset's Rating levels for Confidentiality

CONFIDENTIALITY			
Level	**Title**	**Description**	**Consequence in case of loss of confidentiality**
4	**Confidential Asset**	Asset with a special sensitivity that must be accessed by special authorized staff or services.	***Serious impact***: Damage could affect directly the system, customer or organizations.
3	**Restricted Asset**	Assets that must be accessed only by authorized staff members or services.	***Significant impact:*** the reputation of the system can be harmed.
2	**Internal Asset**	Assets for internal usage in the system, which must be accessed only by internal staff.	***Negligible Impact:*** If the confidentiality is breached, small

			or inconsiderable consequences will happen for the system.
1	**Public Assets**	Assets of the system that can be accessed by anyone or any service.	***Insignificant impact.*** No damages for the System, Customer or Organizations.

Table 2. Asset's Rating levels for Integrity

Integrity			
Level	**Title**	**Description**	**Consequence if there would be an Integrity failure**
4	**High**	The assets must not be compromised by anyone.	**Serious impact**: The consequences could be catastrophic for the system.
3	**Medium**	The assets can be compromised by only service personnel with privileged or extended user rights.	**Significant impact**. The consequences are major and widespread. System errors and services breach persist for a substantial amount of time.
2	**Low**	The assets can be compromised by internal users even if not having any privileged and extended user right.	**Minor Impact**. The consequences are noticeable but workaround can be implemented within the system.
1	**Negligible**	The assets can be compromised by anyone even external users.	**Negligible impact**. Small or inconsiderable consequences which will not have noticeable influence on the system's operation.

Table 3. Valuation levels for Availability

AVAILABILITY			
Level	**Title**	**Description**	**Consequence of Availability deficiency**
4	**Significant**	Unavailability is unacceptable. The asset fails immediately and cannot be re-established by a workaround.	**High impact** on system's operation, which may lead to a complete stop or a main impact on the system. Impacts on the public image of the system and/or of the customer.
3	**Major**	A very short period of unavailability can be accepted during which assets will be unable to provide the intended work.	**Medium impact** affects the system partially and may lead to a delay in the operation of the system.
2	**Minor**	A short period of unavailability can be accepted, assets can be re-established by the implementation of alternative procedures.	**Small impact** on the operation. Small delay with low impact on the operation.
1	**Insignificant**	Unavailability is acceptable. Asset's continuity is not affected.	**Very-small impact** on the operation. No direct delay on the system.

5.4.2 Threat Modeling and Selection: Using RMAT software

Threat modeling and selection step is about preparing a set of appropriate threats and associate them to asset classes and organizing them also into classes. Modeling is meant to prepare the threats selected; and RMAT software can be used in the modeling. RMAT is a software used to prepare elements to the next phases, and in particular to create a threat profile that contains the main threats that are probably to affect the current system. Using RMAT, a TSV (Threat Standard Values) files could be created using a GUI, a TSV file is a representation for threats. Identifying threats for the TSV file is made by associating threats to asset families. The left

panel in figure 18 shows the families and the threats associated to each one, while the right panel shows the threats and the families associated to each one.

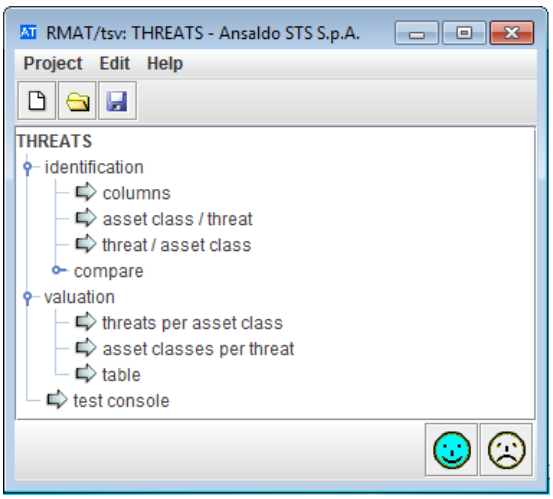

Figure 17. Creating TSV file using RMAT

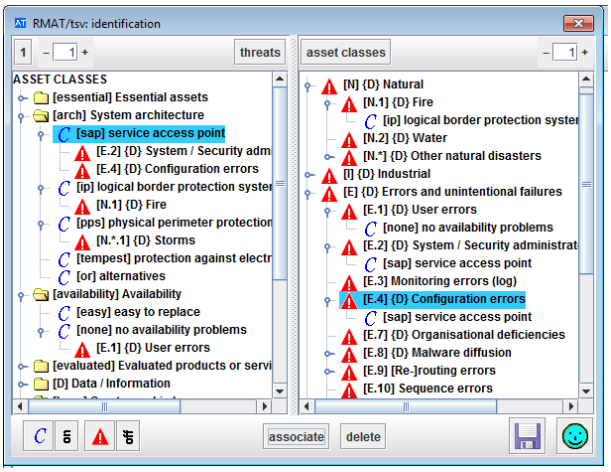

Figure 18. Associating threats to asset classes Using RMAT

The structure of .TSV files that is used to create threat families is:

```
file ::=
        <threat-standard-values>
                { family }0+
        </threat-standard-values>
family ::=
        <family F >
                { threat }0+
        </family>
threat ::=
        <threat Z f [ s ] >
                { set }0+
        </threat>
set ::=
<set D deg />
```

After creating the appropriate set of threat families, next is to use it as input to the risk analysis study.

5.4.3 Conducting Risk management study using MAGERIT method

Following a methodical way in a risk management study is significant in order to obtain an efficient study. The objective of Magerit method is to cover both risk analysis and treatment for a thorough risk management. MAGERIT is suggested by ENISA which is the European

Union Agency for Network and Information Security. MAGERIT is an open methodology for Risk Analysis and Management, developed by the Spanish Ministry of Public Administrations. The purpose of Magerit is directly related to the generalized use of IT systems, communications, and electronic media. This method follows the international concepts as in ISO 31000 and ISO/IEC 27005 [28]. Magerit offers a systematic method for analyzing risks, and helps in describing and planning the appropriate measures for keeping the risks under control. And finally, prepares the organization for the processes of evaluating, auditing, certifying or accrediting, as relevant in each case. On the other hand, PILAR software implements Magerit method is used to perform its steps. Its GUI (graphical user interface) enables the user to execute the MAGERIT method in an understandable and easy way. The tool provides fast calculations and generates a quantity of textual and graphical reports. PILAR software has been funded by the Spanish National Security Agency. It is designed to support the risk management process along long periods, providing incremental analysis as the safeguards improve [29]. PILAR enables the user to create a project, identify the assets for the system under study, and generate threats and safeguards and other functionalities.

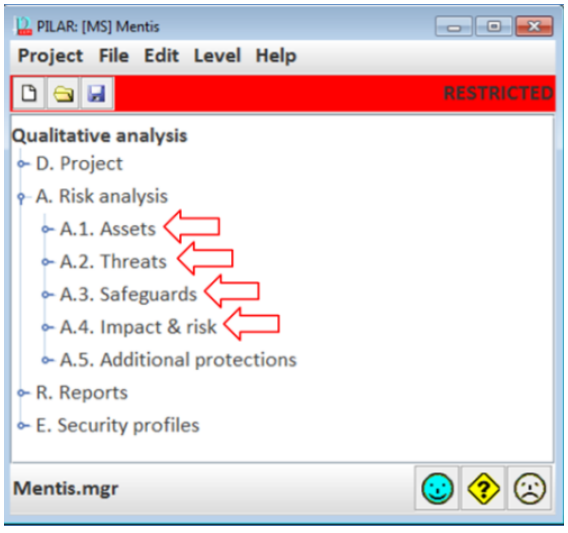

Figure 19. PILAR software: homepage

Furthermore, PILAR can be customized to use TSV files created by RMAT as input for the risk management study, so in this case the threats will be selected based on the model created before in ”Threat Modeling” step.

5.4.4 Safeguard implementation

The safeguard implementation step reflects the “DO” phase of the PDCA, which is putting the chosen decisions in the previous treatment plan into operation. At Hitachi Rail STS, the Defense in depth (DiD) approach is adopted while implementing safeguards, this approach that is based on layering that helps in faster detection and slowing down attacks. In IT environments, DiD is intended to increase the costs of an attack against the organization, by detecting attacks, allowing time to respond to such attacks, and providing layers of defense so that even successful attacks will not fully compromise an organization. A DiD strategy is necessary because of the new security threats and the importance of IT security monitoring of assets.

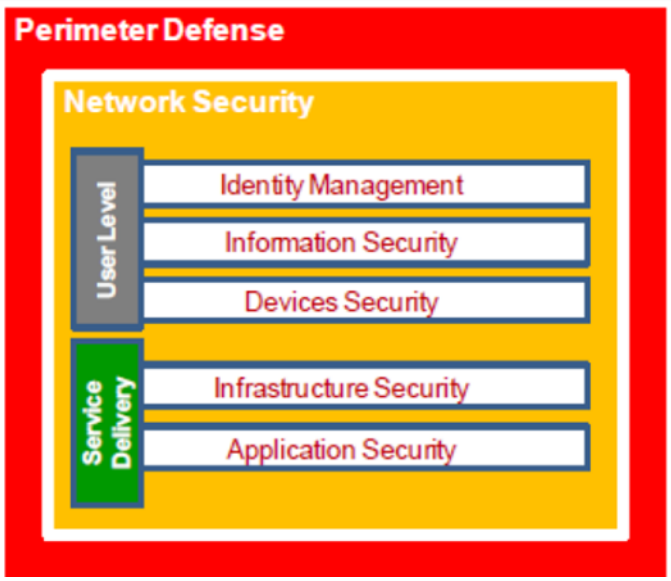

Figure 20. Layering: Defense in Depth

5.4.5 Vulnerability assessment for cyber assets:

The cyber side of a CPS contains various set of assets such as network appliances, servers, software, web applications, databases, etc. At Hitachi Rail STS, vulnerability assessment is applied basically on 3 levels: network, application, and operating system levels.

- **Network Vulnerability assessment:** Network scanners are useful to analyze the network, and hosts on the network to detect vulnerabilities. Nmap (Network Mapper) is a security scanner used on this level to discover hosts and services on a computer network, thus building a "map" of the network. Nmap features include host discovery, port scanning, OS detection, in preparation for auditing, which all help in finding and exploiting vulnerabilities in the network.

- **Web application vulnerability assessment:** Using automated web application and web services vulnerability scanning solutions to apply attack algorithms and determine the existence and relative severity of vulnerabilities. Some dedicated tools employ an extensive arsenal of attack agents designed to detect security flaws in web-based applications. It probes the system with thousands of HTTP requests and evaluates each individual response. This assessment detects vulnerabilities, pinpoints their location in the application, and recommends corrective actions.

- **OS Vulnerability Assessment:** On the level of operating system, what is meant is to apply host vulnerability assessment through scanning specific hosts. This allows the administrators to go beyond testing for known network vulnerabilities, but also

examining more vulnerabilities such as patch levels, check OS configuration, and installed software on computers running operating system.

5.4.6 Compliance:

Compliance can be oriented to internal policies and rules or to external laws and regulations, but in any case it represents a fundamental step in order to maintain the organization control inside its specific regulatory environment**.** PILAR software can be also used to conduct this step by using a security profile that is a description for a list of policies that a system would comply to. It is a collection of safeguards that aim to protect a system. Security profiles may focus on some specific aspects, or may be general. The use of a security profile in a project is basically to check and ensure compliance. It is also possible to create custom security profiles, and some are widely known e.g.: ISO/IEC 27002. PILAR maps security profiles to its safeguards in such a way to estimate to which extent the system is compliant.

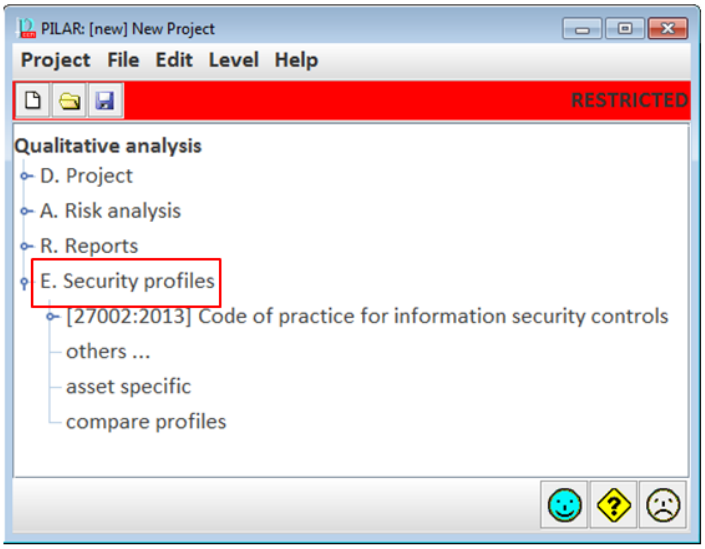

Figure 21. Applying security profiles in the compliance step

After loading a security profile into the project, the goal is to evaluate its set of controls based on the safeguards, to check the compliance of the system to this security profile. The aim is to

focus on the formal compliance to some security standard could be reviewed during the "Audit" phase.

5.4.7 Maintenance and Improvement:

At the end, after executing all the steps of the framework, it is critical to monitor and observe if the taken decisions were effective, and if there is a need for maintenance or improvement or even adding a missed measure. On the other hand, in some situations it could be necessary to reduce the cost of a certain countermeasure. Using PILAR, in the PLAN phase, the "current" stage represents the current state of the system, and "target" stage represents the goal to reach (Fig. 22). However, now in the "ACT" phase, a new target (Fig. 23) will represent the new goal to achieve based on the new observations and analysis done, and putting all safeguards into operation. The system is monitored and a set of investigation and observations is done to apply the refinement in case it requires.

current	target
L1	L1-L3
L2	L3
L1	L1
L1	L1-L3
L1	L1-L3
L2	L2
L3	L4

Figure 22. Safeguards values in PLAN phase

target	new target
L1-L3	L2-L5
L3	L2-L4
L1	L2-L3
L1-L3	L2-L4
L1-L3	L2-L4
L2	L2-L4
L4	L2-L5

Figure 23. New Safeguards values in ACT phase

In recent years, we have seen a growth in the development of various types of Cyber-Physical Systems (CPS). They have brought impacts to almost all aspects of our daily life. Many of such systems are deployed in the critical infrastructure, and so, they are exposed to different types of attacks. In this chapter, a holistic framework is proposed from Hitachi Rail STS's point of view that breaks the restriction to a traditional risk assessment method, and encompasses wider set of procedures which can be followed in any risk mangement study for the CPSs.

6 Chapter six: Guidelines to select an applicable SIEM solution

This chapter discusses how to select a Security Information and Event Management (SIEM) that is most applicable inside a company. It was conducted during the phase Hitachi Rail STS was searching for a suitable SIEM. The work was concluded by proposing an approach that can be followed by organizations trying to do a similar job. This part of the work can be defined as a procedure that falls at the end inside the ISMS.

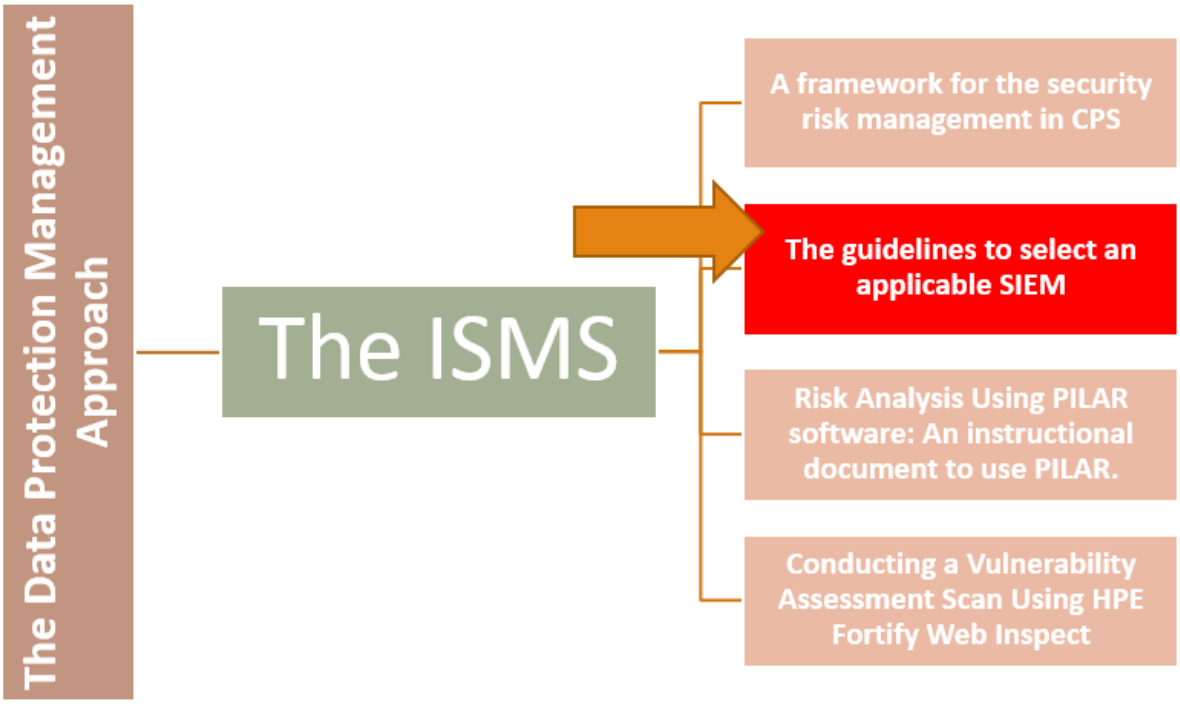

Figure 24. PhD goals (related to the ISMS): The guidelines to select an applicable SIEM

The need for Security Information and Event Management (SIEM) systems increased in the last years. Many companies seek to reinforce their security capabilities to better safeguard against cybersecurity threats, so they adopt multi-layered security strategies that include using a SIEM solution. However, implementing a SIEM solution is not just an installation phase that fits any scenario within any organization; the best SIEM system for an organization may not be suitable at all for another one. An organization should consider other factors along with the

technical side when evaluating a SIEM solution. This chapters describes an approach proposed to aid enterprises, in selecting an applicable SIEM solution.

Information and communication technology (ICT) has made a remarkable impact on the society. Companies nowadays rely on information and communication technology which puts their assets under certain risks especially cyber ones, hence they must be kept under control by means of security countermeasures that generate confidence in the use of these assets [9]. Companies all over the world need to ensure valuable assets, uninterrupted business operation (processes), reliable data and quality of service (QoS) to various groups of users [30]. They need to protect their clients and employees both inside and outside the organization [31]. According to Gartner, by 2020, 30% of global enterprises will have been directly compromised by an independent group of cybercriminals or cyber activists. Moreover, in 60% of network breaches, hackers compromise the network within minutes [32]. On the other hand, companies' IT environment is getting more complex, involving many security appliances that may contribute to security strategy in business processes. Therefore, organizations started to invest on integrating SIEMs (Security Information and Event Management) to improve their security. The term SIEM was introduced by Gartner in 2005. The SIEM system has replaced two types of systems before separated – Security Information Management (SIM) and Security Event Management (SEM) systems [33]. The former provided long-term storage, analysis and reporting, while the latter collected events in real time. Their combination yearned for near real-time analysis, to send notifications and represent information at an operator's console in charge for taking defensive actions. Overall, SIEM system combines SIM and SEM functionalities into one security management system, which collects and correlates relevant data from multiple sources, outputs reports, identifies deviations and takes appropriate actions.

For example, when a potential issue is detected, SIEM might log it as a new information, generate an alert and instruct other security controls to stop any activity progress. Gartner estimates the SIEM market will grow at a compound annual growth rate (CAGR) of 9.5% between 2016 and 2022, and the worldwide spending on SIEM will reach 3.72 billion dollars [34].

From an organization perspective, the challenge is not just about selecting any SIEM solution but implementing the right solution that fits better within company structure and is aligned with the existing threats landscape. In addition, it must be flexible enough to be easily adapted to meet any changes thereafter. Security and risk management (SRM) leaders evaluating SIEM solutions must understand their use cases and then define specific requirements in conjunction with applicable stakeholders and company strategy in general [35].

On the other hand, organizations must require a structured approach for managing their challenges. This will ensure that there are agreed objectives, good management controls in place and effective monitoring of performance to keep on track and avoid unexpected outcomes. Therefore, this chapter proposes not just technological but pragmatic approach to support companies that are seeking to adopt SIEM systems into their environments, suggesting suitable answers to preferred requirements that are believed to be valuable prerequisites a SIEM system should have. The aim of the proposed approach is to advice a pre-installation strategy, a way to evaluate functional components that a SIEM should comprise in terms of both technical and organizational requirements, and to suggest criteria to judge SIEM systems using an evaluation process composed of quantitative and qualitative methods.

However, and because of the complexity, precision and thoroughness required to apply our whole approach, it is mainly dedicated to large enterprises, which include wide variety of broad

and specific skills and several specialists to manage certain applications or parts of the IT infrastructure, and most of them comprise a dedicated department to manage information security. Therefore, this approach can be followed by those bigger enterprises that in general tend to manage their work in a very structured manner, they need to assert successful management and performance, and this is our goal, to aid in following a thorough approach for the issue. This work represents the first and primary phase in the procedure of choosing a SIEM or a set of qualified SIEMs, however, it has to be followed by a testing phase that enable the customer to check the solution directly after the installation.

6.1 Backgrounds and related works

6.1.1 SIEM system: Definitions

There is a plethora of features regarding SIEM systems, which are developed differently by each vendor. In general, SIEM collects, normalizes and aggregates event data produced by security devices, network infrastructure, systems, and applications. Event data is combined with contextual information about users, assets, threats and vulnerabilities. SIEM systems could be agentless and agent based [36] [37], or even hybrid (using both agent and agentless) and may adopt new technologies such as HEC (Http Event Collection). Agentless means that the log-source transmits its logs to the SIEM, or an intermediate logging server involved, such as a syslog server; while agent-based means that an agent is installed on a source-log to gather security events from the endpoint itself. Today, most SIEM systems work deploying multiple collection agents (collectors) in a hierarchical manner. Log collectors forward events to a centralized management console, which performs inspections and flags anomalies [31]. Then after collection, the data should be normalized so it can be correlated and analysed. Another feature is the pre-filtering that is related to processing centre, some systems use a pre-processing mechanism at the edge collectors, with only certain events being passed through to a centralized

management node. In this way, the volume of information being communicated and stored can be reduced.

SIEM technology provides near real-time correlation of events for security monitoring, query and analytics for historical analysis and other support for incident investigation, compliance reporting, and alerting [38]. According to Gartner, by 2020, 75% of all SIEM solutions will use big data technologies at their core, along with machine learning, to improve threat detection and response capabilities [35]. In short, there are so many SIEM systems in the market created by skilled and expert security vendors with their own features, however, selecting the suitable SIEM is not a trivial task anymore: it is not simply about installing the most powerful one, for instance, a very powerful SIEM may be too complex to apply in some cases.

6.1.2 Related works

Several studies were conducted in the field of "How to select a SIEM system". Gartner reports are an excellent example where they present a detailed evaluation of the current SIEM products based on many characteristics such as sales execution, pricing, costumer experience, marketing message evaluation against the understanding of customer needs [35] [38] [39]. In [38], the authors examined different SIEM products that are the leaders of SIEM technology, they focused on technical requirements and showed strengths and cautions for each vendor and at the end they defined evaluation criteria from an ability-to-execute point of view. According to [35], the analysts defined a list of critical capabilities that a SIEM should include and suggested three different and general use cases: Basic Security Monitoring, Complex Security Monitoring, Advanced Threat Defence. They used an evaluation criterion to evaluate the most powerful SIEM products available, which is based on the defined critical capabilities and the

three suggested scenarios. In [40], SANS Institute provided environment-specific criteria to benchmark SIEM solutions, where organizations should consider factors like Events-Per-Second (EPS), considering the number of employees in each sub-net, number of databases, and the ability to store and analyze these events, in order to evaluate and even design a SIEM system. In [41], Nabil et al. proposed an after-installation evaluation approach: such an approach may be time-consuming for companies, and it should be preceded by a PRE-installation evaluation approach that qualifies and select the applicable SIEMs from the plethora of solutions available before installing them.

The common point between the above-defined related-work is that all of them are product driven, where the evaluation of the solution is for the product as a technical solution. However, our approach is customer driven, where the selection phase is not only based on the technical features of the product but also subject to pre-defined customer's needs.

6.2 Aspects to be addressed before adopting a SIEM solution

Before starting the procedure of choosing a SIEM, it is essential to consider some general aspects that could influence the technical and the organizational evaluation of the solution. Companies must understand completely the problems they are trying to solve, considering aspects such as company type, its assets, what to secure, internal policies and external regulations.

To select an appropriate SIEM, the customer should prepare a list of requirements to describe the needs sought in the SIEM, usually described in a request-for-proposal (RFP) document. However, before defining those requirements, some general aspects that affect the SIEM selection should be examined, and SRM leaders must include at least the following considerations as a first step:

- **The company:**

Companies may differ by vertical, size, location, and other factors that may affect SRM leaders' decisions. Each company has its own information system and architecture, some companies could be geographically distributed, require high availability services and may generate enormous amount of logs with different formats.

- **Prioritizing assets and risks:**

It is necessary to conduct an overall study to identify and evaluate the most critical assets and their corresponding risks in order to specify the most important targets of the monitoring and defensive activities of the SIEM system. The goal is to identify odds and costs if something wretched happened. Prioritizing risks helps in selecting what logs are more important, to give them high priority when configuring the SIEM for more efficient correlation and reporting.

- **Compliance, regulations and forensics capability:**

The company might be impacted by internal or external regulations. Compliance could be like conforming to international standards, or internal policies. For example, in some cases, it may be necessary to keep log-collectors on site to secure data in the place of origin, based on a specific state or country regulation, hence the entire deployment will be affected with such regulations, and so they should be considered. On the other hand, forensics should be addressed from two sides, the first is about ensuring, once collected, the logs were not altered and their integrity were preserved, the other side of forensics is that the SIEM should support incident management and investigation activities.

- **Security Operation Center:**

One essential aspect is about who will own, maintain and operate this new technology. SOC (Security Operation Center) is a centralized unit that deals with security issues on an organizational level made up of a team primarily composed of security analysts (and operators) organized to detect, analyse, respond to, report on, and prevent security incidents in order to minimize risks [37]. This choice is part of a more general view oriented to the integration into company IT environment and effort needed to maintain such a system during license lifetime. And so, in the selection of a SIEM, the responsible should be aware of who will operate the SIEM, and if they can operate the selected SIEM platform or by considering the training costs.

- **Human and technological aspects:**

Finally, an important aspect that should be also considered is to gather feedback from inside the organization about the IT resources and capabilities to support a specific SIEM product and potentially overlapping technologies in order to release IT means and personnel and skills to be allocated in this activity. Moreover, the organization should be aware about how the selected platform must be easily integrated in the technology environment of the company.

6.3 A SIEM selection approach: Requirements and Evaluation

What characterizes this work is that it proposes an overall approach for the problem of selecting the applicable SIEM solution, and searching the previous work will show how few, if no similar comprehensive approaches were proposed. Some of the work done focused just on the technical requirements without addressing the organizational ones, and other aspects. Others did not consider the problem of applicability or integration in the environment. This approach, unlike others, is customer driven which means that customer needs are taken into account when

following the whole approach, specifically when defining the requirements and then evaluating the suppliers' solutions.

Saving time, using a systematic-organized strategy for decision-making and balancing costs to needs are the main advantages of adopting such an approach. This approach starts by suggesting the requirements (technical and organizational) that should be addressed in a SIEM solution in a systematic way , and then proposes a methodology for evaluating SIEM solutions that measures the compliance and applicability of any SIEM solution using quantitative and qualitative methods. This evaluation methodology is split into two phases:

(1) Quantifying each requirement of the received SIEM solution using a quantitative based method;

(2) Measuring the applicability of the solution using a qualitative based method after defining a list of indicators that enables the evaluation of this applicability.

The goal is to select the appropriate SIEM that fits best in company's environment and resources; however, we stress that a final installation-testing phase must be accomplished with the suppliers to make sure about the compliance of the selected solution to the needs addressed.

6.3.1 Technical and organizational SIEM requirements:

Defining requirements is an important task; it helps the companies to define their needs, be aware of any shortage, and aids them in the evaluation phase. Information security and risk management leaders responsible for security operations should focus their evaluation on the critical capabilities that align with their use cases, requirements, and current and future IT environments, e.g. on-premises versus cloud-based services [35].

This section groups the requirements needed to adopt a SIEM platform covering in detail the "*mandatory"* and *"nice-to-have"* requirements, as listed in Table 4.

Those requirements represent the needs of the customer, they cover all the features and services that the supplier should include when proposing a SIEM solution.

Table 4. SIEM requirements

Section	Type	Requirement
Platform	**Mandatory**	1. Log Management System capability 2. Supporting an extended set of log sources 3. Customization of parsers/connectors 4. Method for retrieving events/flows/logs 5. Specification of the method for retrieving events/flows/logs 6. Hierarchical and modular/scalable architecture 7. Time-zones management 8. Platform computing capacity 9. Platform storage capacity 10. Installation model 11. High Availability/caching options 12. Availability of both default and customizable correlation rules 13. Dashboard features: ability to quickly prioritize response and analysis 14. Customizable and compliance reports 15. Alerting capabilities 16. Technical documentation and online help 17. Ability of Monitoring the platform 18. Secure Software 19. Context enrichment based on collected logs

		20. Support for collection of real-time and deferred logs
	Nice to have	21. Multi-tenant capabilities (views) 22. Anonymization of logs 23. Support MITRE ATT&CK correlation matrix
Operations	**Mandatory**	1. Role-based access control 2. Accounting: log events done by operators 3. Web interface for day-by-day operation
	Nice to have	4. Customizable time-zones for the GUI
Integration	**Mandatory**	1. Active Directory integration for administrative management
	Nice to have	2. Integration with asset management tools 3. Case Management and trouble-ticketing activities tracking 4. Trouble ticketing module 5. Integration with vulnerability management tools
Advanced features	**Nice to have**	1. Threat Intelligence analysis tools support 2. Support for forensics analysis activities 3. Analytics support 4. Automatic response capabilities

Licensing and support	**Mandatory**	1. Specification of the preferred License type 2. Licensing restrictions 3. Specification of the project Roadmap 4. Delayed license activation 5. Technical assistance support and professional services 6. Training provided

Requirements are divided into 5 sections: platform, operations, integration, advanced features and licensing-support services.

a. **Platform:** Describes the technical requirements needed in the platform.

b. **Operations:** Groups the requirements needed to manage the solution.

c. **Integration:** This section groups the requirements needed to integrate the SIEM solution into the Company's information system.

d. **Advanced features:** This section describes the advanced features, they could be considered as nice-to-have requirements.

e. **Licensing and support:** This section lists and describes the licensing and support services requirement.

6.3.2 Measuring the compliance and applicability of a SIEM: An evaluation process

Evaluation is the structured interpretation and giving of meaning to predicted or actual impacts of proposals or results. It looks at original objectives, and at what is either predicted or what was accomplished and how it was accomplished [42]. It can assist an organization to assess and help in decision-making; or to ascertain the degree of achievement or value about the aim and objectives and results of any action. Evaluation is methodologically diverse; two types of

methods may be qualitative or quantitative. Quantitative methods are distinguished by emphasis on numbers, measurement, experimental design, and statistical analysis [43], and hopes the numbers will yield an unbiased result that can be generalized to some larger population. However, qualitative methods evaluate other parameters such the success and the eligibility of a product in a specific environment using non-numerical (textual forms) data to assess the eligibility and reliability of adopting the solution, such as the use of internal discussions, interviews, comparisons to provide feedbacks, etc. Both quantitative and qualitative evaluation methods have their benefits, quantitative evaluation can help remove human bias, thus more accurate. However, qualitative evaluations may also involve truths, but these truths are harder to get at, and evaluators may not always agree. In our approach, an evaluation process is proposed; it is applied after receiving the description of the SIEM solution from suppliers. It is divided into two methods: quantitative and qualitative. The first method *"Requirements-based Evaluation"* is the quantitative side of the evaluation process; it evaluates the degree of compliance for each requirement of the received SIEM solution using numerical values and mathematical operations. This method is applied to the SIEM solutions that might be adopted and used to obtain a list of qualified ones as an output. After that, the output of this method is then provided as input to the second method .

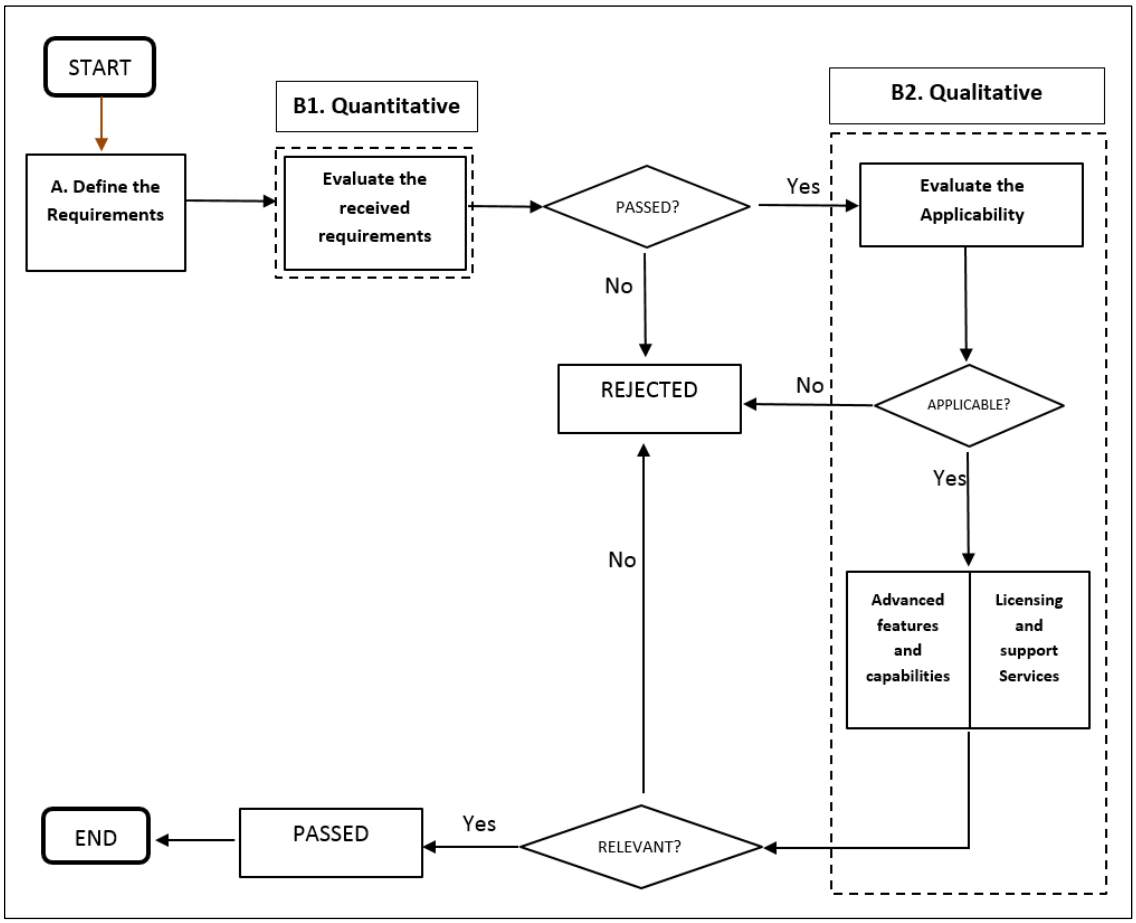

Figure 25. How to apply the overall approach

On the other hand, the second method *"Applicability Evaluation"* represents the qualitative side of this evaluation process, it focuses on the observations, interpretation and the opinion of the concerned parties, rather than going into measuring the value of each requirement of the received SIEM solution. Both methods are complementary in the evaluation process, using both helps in getting a deeper understanding and obtaining a precise and flexible evaluation. The quantitative side represents the accuracy that evaluates and qualify a set of SIEMs, however the qualitative side represents the flexibility, where the evaluators add their analysis, opinion and understanding to compare the qualified solutions and finally to select the most applicable one. Figure 25 shows how to apply the approach starting by defining the requirements that the customer seeks in their SIEM product, then evaluating the received ones using the proposed evaluation process described in the next section.

6.3.2.1 Compliance measurement: A quantitative requirement-based evaluation method

The first evaluation method measures the degree of compliance of each requirement in the SIEM solution that might be adopted, it is a first step evaluation. After receiving the tenders from diverse suppliers proposing a SIEM solution, security and risk management (SRM) leaders evaluate each SIEM solution separately, where each requirement is evaluated to get a total score for the whole solution. Two different parameters are assigned to each and they are the requirement *value* (V) and the *weight* (W). The requirement value is the grade assigned to evaluate the answer-to-requirement in the under-evaluation SIEM solution, while the weight represents the importance of current requirement in the solution from the user point of view, and is assigned initially when the customer defines his requirements, for example, a mandatory requirement has a high value compared to nice-to-have ones. And then, a *score* (S) is calculated for each requirement ($S = V*W$), and after that a total for each requirement section is computed. Finally, the total score is obtained by adding all the totals corresponding the requirements family sections. Table 5 is an example to use in applying this evaluation method.

For a better evaluation, a scale is suggested to represent the requirement value (V). This scale is used to differentiate between an insufficient, good, very good, and excellent requirement proposed by the supplier, by translating the level of compliance of the under-evaluation requirement into a numerical value. A non-linear growth scale is suggested to be used because of its ability to differentiate between the values using the high growth rate.

At the end, evaluators may define a passing grade to use to select the qualified "under-evaluation SIEM solutions, so they can directly reject a solution with lower total score. The output of this method should be several accepted SIEM solutions which all complied the defined

requirements, but at the end one solution should be adopted, and this is the role of the second evaluation method, which examines the applicability.

Table 5. Requirement-based Evaluation

SIEM X	**Requirement**	**Requirement Value(V)**	**Weight (W)**	**Score(s)**	**Total**
PLATFORM	a_1 . . a_{23}	$V(a_1)$ . . $V(a_{23})$	$W(a_1)$ . . $W(a_{23})$	$V(a_1)*W(a_1)$ . . $Va_{23}*Wa_{23}$	$\sum_{i=1}^{23} V(ai) * W(ai)$
OPERATIONS	b_1 . . b_4	$V(b_1)$ . . $V(b_4)$	$W(b_1)$ . . $W(b_4)$	$V(b_1)*W(b_1)$ . . $V(b_4)*W(b_4)$	$\sum_{i=1}^{4} V(bi) * W(bi)$
INTEGRATIONS	c_1 . . c_5	$V(c_1)$ . . $V(c_5)$	$W(c_1)$ . . $W(c_5)$	$V(c_1)* W(c_1)$ . . $V(c_5)*W(c_5)$	$\sum_{i=1}^{5} V(ci) * W(ci)$
ADVANCED FEATURES	d_1 . . d_4	Vd_1 . . Vd_4	Wd_1 . . Wd_4	. . .	$\sum_{i=1}^{4} V(di) * W(di)$

LICENSING AND SUPPORT	e_1 . . e_6	Ve_1 . . Ve_6	. . .	. . . .	$\sum_{i=1}^{6} v(ei) * w(ei)$
TOTAL SCORE					∑: Total

6.3.2.2 Applicability evaluation: A qualitative method

Security and risk management leaders increasingly seek SIEM solutions with capabilities that support early targeted attack detection and response. Users must balance advanced SIEM capabilities with the resources needed to run and tune the solution [38]. The best SIEM system for an organization may not be suitable at all for another. Other variations should be considered along with the technical side when evaluating a SIEM solution. Therefore, the qualitative side of the approach takes place; it is about examining the whole solution in terms of applicability rather than measuring mathematically the value of each requirement. The highest-grade solution is not always the choice, it may have powerful features, but too complex to install, or even too expensive. This method aims to evaluate the qualified SIEM solutions in a high-level manner. It does not aim to evaluate technically each solution, however, to examine the applicability of them. In such method, a unified scale is followed, and a set of *Indicators* is used to evaluate and compare, without going deeply into technical details as in the requirement-based evaluation. *Weight*, *evaluation*, *notes* are other parameters used, and described next.

INDICATORS:

Indicators are grouped into families, and have different weights (high, medium, and low), where some are key factors in the selection process more than others are. They help adequately evaluate technology solution vendors and then decide which solution fits better.

a. The PLATFORM

The "platform" family-of-indicators tries to assess the applicability of the proposed solution. This section encompasses the following list of indicators and should be considered important.

- **Compliance:** the compliance indicator represents to what extent is the proposed solution compliant. In other words, it evaluates the compliance of the mandatory requirements or the existence of non-compliant requirements, taking into account the restrictions, constraints, regulations or policies that prevent the implementation of such solution: e.g., kind of the solution proposed: software/hardware.
- **Quality of services:** a general evaluation for the quality of the services, capabilities that the solution offers.
- **Robust Architecture:** evaluates the proposed architecture of the solution, the deployment, and if this architecture preserves a high availability.
- **Scalability:** evaluates the ease and the ability of the solution to grow, in terms of adding additional features in the future, e.g.: adding additional licenses, etc.
- **Complexity of the solution:** this indicator evaluates the applicability of the platform (platform kind, number of nodes, etc.), ease of deployment, level of integration,

relevance. For instance, the user does not need to modify or develop something hard to integrate the solution.

- **Clearness:** evaluates the clearness of description of the SIEM solution (e.g. Does the received RFP include a complete description or is there something ambiguous?).

b. Licensing and support Services

- **Duration**: Evaluation of the planned duration by the supplier to install the solution (e.g. Does it have a clear roadmap).
- **Licensing**: Evaluates the type of the licensing offered by the supplier (license or other purchase options, e.g. leasing), and evaluates if the activation starts after the end of the acceptance tests in which all the project requirements will be met.
- **The support**: Evaluation for the availability of the technical support (e.g. 7 days/week and 24h/24h).
- **Training**: evaluates the training level provided.

c. Advanced features:

- **Support additional features:** Evaluation of the available advanced features or additional ones.
- **Integration with third parties:** Evaluates how much the solution can be integrated with 3rd party tools, or just restricted or limited.

d. Other Indicators

- **Skill of the supplier/vendor**: Examines if the supplier or vendor has the expertise in this field, and the services that it offers.

- **The price:** Represents an important indicator in the selection process and a cost-effective option should be selected.

WEIGHT:

It corresponds to the weight of the indicator that will be evaluated; the weight in terms of its relative importance in the whole solution, weight could take different values such as high, medium or low. It is up to the evaluator to assign those values based on their own needs and addressing related aspects.

EVALUATION:

An evaluation is defined for each indicator. The evaluation is carried out based on the eligibility and reliability. Values could be insufficient, good, very good, and excellent.

NOTES:

It could be the team's general conclusion drawn up based on the described solution in each tender. It represent additional notes, which the evaluator considers that should be highlighted.

Table 6. Applicability Evaluation for each SIEM solution

SIEM X	**Indicators**	**Importance or weight**	**Evaluation**	**NOTES**
Applicability	*Level of Compliance*	……	……	……
	Complexity	……	……	……
	Quality of services/ capabilities	……	……	……
	Robust Architecture	……	……	……
	Scalability	……	……	……

	Clearness/ complete description	……	……	……
Licensing and support services	*Installation duration/ clearness of road map*	……	……	……
	Licensing	……	……	……
	Support	……	……	……
	Training	……	……	……
Advanced Features	*Additional features*	……	……	……
	Integration with third parties	……	……	……
Other indicators	*Expertise/Skill of Vendor/ Supplier*	……	……	……
	Price	……	……	……
OVERALL RESULT				ACCEPTED or REJECTED

Table 6 shows a model (table-form) to use in applying this proposed method in an easy way. At the end, the suggested indicators may intersect or overlap, so the evaluators may merge some together or even split others. On the other hand, each organization may assign a different weight for those indicators based on the requirements they need, policies, regulations, etc.

6.4 Case study: Applying the approach

The approach presented in this work was applied at Hitachi Rail STS Company [24]. To select a SIEM solution, the Company believes that a structured and systematic procedure should be followed, in an organized manner, which is the way Hitachi Rail STS tends to manage its challenges, i.e. ensuring that there are agreed objectives, good management controls in place to keep on track and avoid unexpected outcomes. Another main reason for selecting and

adopting such an approach is to provide enhanced compliance and requirements coverage in bids.

6.4.1 Creating a Request-For-Proposal (RFP): Specifying SIEM requirements

Specifying requirements is the first step while applying this approach. Therefore, a request-for-proposal (RPF) document (containing all the requirements that are believed to be mandatory or nice-to-have in the SIEM quest) was prepared and inspired by the SIEM requirement section of this approach, extending and describing briefly each requirement, to make the suppliers aware of the features that their solution should have to fit the customer needs. Following the approach proposed, requirements are divided into 5 sections: platform, operations, integration, advanced features and licensing-support services. The RFP sent to the suppliers:

a. Platform. Describes the technical requirements needed in the platform.

1. **Log Management System capability**: The technical solution must address the collection, hashing, normalization, indexing, compression plus archiving, retention, (and all usual Log Management Systems' features) of events and log files along with aggregation, correlation, analysis, reporting and alerting. The platform should be easily integrated with the most widespread log management or SIEM solutions.
2. **SIEM platform kind:** The supplier should provide details about the kind of SIEM platform available (for example hardware appliances or virtual appliances, or software only), where hardware is preferred in our case.
3. **Supporting an extended set of log sources:** The platform must be able to parse with native support the most widespread log sources. A list of the mandatory ones must be listed.

4. **Customization of parsers/connectors:** The platform should be able to support the creation of a library of customized parsers/connectors.
5. **Method for retrieving events/flows/logs:** Specification of the method used for retrieving events/flows/logs (by agent/agentless support): the customer must be aware about the relevant method that fits the case, and the supplier should describe his proposed one clearing why the selected method is better.
6. **Hierarchical and modular/scalable architecture:** The architecture should be scalable featured by unlocking license or adding modules, without the need of replacement and reconfiguration. Additional value if regional-based hierarchy is supported, with local collection and caching at main nodes and intelligent correlation at a central post.
7. **Time-zones management:** The SIEM architecture must support many different time-zone management capabilities, even up to providing time-zones when capturing log files which have none.
8. **Platform computing capacity:** The calculation to determine the appropriate number of sustained EPS and the EPS peak value proposed, should all be stated. Such parameter must be conveniently increased to support events storm in peak situations.
9. **Platform storage capacity**: The platform must be able to store the events/logs for an agreed period of time (e.g.: in months) for a quick indexed access and for a long-term storage.
10. **Installation model**: The supplier should specify what installation model is available (e.g. on premise, private cloud or managed option).

11. **High Availability/caching options**: The redundancy/caching options should be available to avoid event/log file transfer losses in case of the distributed installation. Additional value if load balancing is possible among remote nodes while sustaining the same number of EPS.
12. **Availability of both default and customizable correlation rules**: The platform should include a set of standard correlation rules scenarios, and must be able to design further correlations rules.
13. **Dashboard features**: The dashboard should be able to quickly prioritize response and analysis (e.g.: multiple drill-down, elastic search, customizable dashboards per user, default views)
14. **Customizable and compliance reports**: The supplier should provide details about the availability out of the box of compliance reports (e.g., ISO27001, SOX, NERC, Top 20 CSC, Privacy, etc.) and the generation of customizable ones.
15. **Alerting capabilities**: Capability of triggering alerts, e.g. sending a notification message or email, and so to respond to incidents.
16. **Technical documentation and online help**: Availability of technical documentation, both with offline and online help (E.G. available wizards, recipes, etc).
17. **Monitoring***:* The SIEM platform should be monitored using any standard protocol (e.g.: SNMP) so the it can be added in the Company's monitoring platform.
18. **Secure Software: T**he supplier should state the SIEM platform operating system and version along with the "secure by design" techniques adopted.

19. **Context enrichment based on collected logs:** Availability of correlation rules to gather and merge information from the different log sources, to be able to provide all the info of the array [Mac, IP, hostname, username] whenever available somewhere in the logs.
20. **Support for collection of real-time and deferred logs:** The supplier should provide details about the support, normalization and indexing of logs retrieved in real-time and deferred way (e.g. sent by batch jobs).
21. **Multi-tenant capabilities (views):** The supplier should provide details about the SIEM solution capability to display some views based on connector grouping (e.g., connectors associated to different geographical entities or based upon management responsibilities like systems, DBs, network or security devices).
22. **Anonymization of logs, e.g. for GDPR compliance:** Provide details about anonymization of logs (e.g., by at least masquerading privacy-related info to some profiles of users).
23. **Support MITRE ATT&CK correlation matrix:** Provide details about the support of TTP use detection of MITRE ATT&CK matrix.

b. **Operations.** Groups the requirements needed to manage the solution.

1. **Role-based access control:** The platform has to implement a role based access control mechanism suitable by the configuration of multiple user profiles owning different privileges to implement the accountability and separation of duties principles.
2. **Accounting:** The SIEM platform must have an audit log facility in order to track the activity relevant from the security perspective performed by operators.

3. **Web interface for day-by-day operation:** The SIEM platform interface used by users for daily analysis has to be web-based.
4. **Customizable time-zones for the GUI:** The interface must allow the user to choose in which time zone all the data must be displayed.

c. **Integration.** This section groups the requirements needed to integrate the SIEM solution into the Company's information system.

1. **Active Directory integration for administrative management:** The platform access must be granted only to qualified users authenticated and authorized via the Company's Active Directory database.
2. **Integration with asset management tools:** The ability to integrate the solution with asset management standard tools (E.g. Configuration Management Data Base).
3. **Case Management and trouble-ticketing activities tracking:** The ability to manage incident handling issues and the conditional support of standard IT trouble ticketing systems (workflows, prioritization, KB email exchange).
4. **Trouble ticketing module*:*** The supplier should provide details in case of the availability of a trouble ticketing system with the SIEM.
5. **Integration with vulnerability management tools:** A nice feature is the ability to integrate with vulnerability management tools.

d. **Advanced features.** This section describes the advanced features, they could be considered as *nice-to-have* requirements.

1. **Threat Intelligence analysis tools support:** Availability of threat analysis tools is a plus if already available by the vendor and by using standard formats for exchange such as: STIX, TAXII, IoC, other standard formats.
2. **Support for forensics analysis activities:** Additional value if forensics analysis activities are available (file integrity monitoring, pcaps, NetFlow, evidence acquisition).
3. **Analytics support:** Additional value if there is a support for anomaly detection, use and entity behavioral profiling.
4. **Automatic response capabilities:** Additional value if there is a support for automatic response capabilities (e.g. SOAR: security orchestration automatic response).

e. **Licensing and support:** This section lists and describes the licensing and support services requirement.

1. **Preferred License type**: The supplier should specify if the SIEM solution would be available only by license or also in other purchase option, e.g. such as leasing.
2. **Licensing restrictions**: State any license limitations, for example what happens in case of exceeding the limits mentioned in license.
3. **Project Roadmap**: Describe the tasks involved with the project of the SIEM platform installation and configuration and the corresponding timeframes.
4. **Delayed license activation**: The license activation should start only after the end of the acceptance tests in which all the project requirements will be met. The customer must ask for an acceptance-testing period to verify that the solution complies with the received description.

5. **Technical assistance support and professional services:** The supplier should include a description of the including technical support and professional services provided/available by vendor/system integrator.
6. **Training provided:** A description of training package should be provided.

6.4.2 Evaluating the received SIEM solutions

Different solutions were proposed by a set of suppliers according to their expertise in the field and using the most powerful SIEM products available in the market nowadays. However, in this case study, only four solutions are selected to apply the proposed approach, which are believed to be enough for showing and verifying the efficiency and effectiveness of our approach. But, for the purpose of not being subjective or promoting a solution, the study will not mention any supplier or SIEM product name, and instead it will anonymize their names and use the following notation for them:

(1) **Supplier 1 using product X**
(2) **Supplier 2 using Product Y**
(3) **Supplier 3 using Product Z**
(4) **Supplier 4 using Product T**

Where the "supplier" is the one who provided the whole SIEM solution (product, architecture, installation, licenses, training, etc.) and the "product" is the name of the innovator who created the SIEM product.

After receiving the tenders from the suppliers, describing their overall solutions, the next step in the approach is to apply the evaluation phase. Evaluation is divided into two methods, *requirements-based* (quantitative) and *applicability-based* (qualitative), where the first aims to qualify a set of the best in terms of matching and complying the requirements specified by the customer, and the latter aims to select only one solution that best fits the company.

IT security and risk management leaders responsible for security operations should focus their evaluation on the critical capabilities of the SIEM that align with their use cases, requirements, and current and future IT environments [39], therefore it is not just about selecting the best SIEM in the market.

And so, the evaluation done next is based on the whole solution offered, which includes the proposed architecture (topology for installation and deployment), kind of the platform, complexity, technical features, licensing, etc…, and it is not only an evaluation for the tool or the vendor that develops it. Again, the evaluation applied using this approach is from a customer point of view, ensuring the selection of the most appropriate and applicable solution based on the needs and context.

6.4.2.1 Requirement-based evaluation: Using a quantitative method

According to the received tenders, only four solutions were selected to demonstrate the evaluation of our approach, which we believe are enough to show the effectiveness of such an approach, knowing that they used the most powerful SIEMs existing in the market nowadays that are developed by well-known, skilful and expert vendors.

As described in section 6.3.2.1, each requirement is assigned two different parameters, which are the *weight* (W) and the *requirement value* (V):

- The *weight* (W) represents the importance of current requirement in the solution from the user point of view, and is assigned initially when the customer defines his requirements, for example, a *mandatory* requirement has a high value compared to *nice-to-have* ones.
- The *requirement value* (V) is the grade assigned to evaluate the answer of the supplier to the current requirement in the under-evaluation SIEM solution.

Then, a *score* (S) is calculated for each requirement (*S=W*V*), and after that a total for each requirement section is computed, and finally a total score is computed for the whole solution.

The approach also suggested to use a non-linear growth scale because of its ability to differentiate between the values using the high growth rate, and so requirement value will get a non-linear value as suggested in the approach and will vary between 0, 1, 2 and 4 corresponding a zero, low, medium and high requirement evaluation. On the other hand, the weight will vary between zero and one, where one represents a critical requirement.

Therefore, to evaluate those solutions, an quantitative evaluation is used to show the *requirement-based* evaluation. Tables 7, 8, 9, 10 present the quantitative evaluation of the four received SIEM solutions.

SUPPLIER 1, PRODUCT X:

Supplier 1 offered a solution based on a product X, the solution composed of three layers:

(a) Collection layer that uses collection nodes to collect the logs, and to parse and normalize.

(b) Processing layer for data storage and correlations based on specific rules.

(c) Analysis layer for reporting, and alerts or offense investigation.

In addition, the supplier offered a solution that can be deployed using two alternative architecture for deployment:

(a) All-in-One deployment, where all layers are within a single node. Multiple nodes (dedicated appliances HW+SW) will be used based on the need,

(b) Distributed deployment, that consists of multiple nodes, composed of multiple nodes (dedicated appliances HW+SW).

The solution covers most of the requirements, e.g. indexing, normalization, compression, and hashing for tamper proof, and it is able to parse a lot of sources and supports also custom parsing.

Moreover, it is able to store events/logs locally for a long period. In addition, based on the proposed architecture of deployment, high availability is preserved.

Table 7 Evaluation of the SIEM solution by Supplier 1

Product X Supplier 1	**Requirement**	**Weight (W)**	**Value (V)**	**Score(S)**	**Total**
PLATFORM	a_1	1.0	4	4	43.8
	a_2	1.0	2	2	
	a_3	0.9	2	1.8	
	a_4	0.8	2	1.6	
	a_5	1.0	4	4	
	a_6	1.0	2	2	
	a_7	1.0	2	2	
	a_8	1.0	2	2	
	a_9	1.0	2	2	
	a_{10}	1.0	4	4	
	a_{11}	1.0	2	2	
	a_{12}	0.8	2	1.6	
	a_{13}	0.8	1	0.8	
	a_{14}	0.8	1	0.8	
	a_{15}	1.0	1	1	
	a_{16}	0.8	2	1.6	
	a_{17}	0.8	1	0.8	
	a_{18}	1.0	2	2	
	a_{19}	1.0	4	4	
	a_{20}	0.9	2	1.8	
	a_{21}	0.6	1	0.6	
	a_{22}	0.7	2	1.4	

	a_{23}	0.5	0	0	
OPERATIONS	b_1	1.0	1	1	
	b_2	1.0	1	1	3.8
	b_3	0.9	2	1.8	
	b_4	0.8	0	0	
INTERGRATIONS	c_1	0.8	1	0.8	2.1
	c_2	0.5	0	0	
	c_3	0.8	0	0	
	c_4	0.8	1	0.8	
	c_5	0.5	1	0.5	
ADVANCED FEATURES	d_1	0.7	2	1.4	3.7
	d_2	0.6	1	0.6	
	d_3	0.6	2	1.2	
	d_4	0.5	1	0.5	
LICENSING AND SUPPORT	e_1	1.0	2	2	12.6
	e_2	1.0	1	1	
	e_3	1.0	2	2	
	e_4	0.8	2	1.6	
	e_5	1.0	2	2	
	e_6	1.0	4	4	
				TOTAL SCORE	66

SUPPLIER 2, PRODUCT X:

The solution proposed by supplier two uses product Y. It has a modular and distributed architecture, and is composed of two layers:

(a) A layer that groups collection and standardization, it can receive logs of any type of data source within the network without the need to install agent designed for each specific platform.

(b) A layer for correlation on the main appliance.

The solution uses a hardware-based appliance (On premise or in cloud) that is able to support continuous traffic.

However, the solution lacks a lot of the main requirements such as indexing, hashing, normalization, customizable dashboard per user, and for compliance reports, etc…

Table 8. Evaluation of the SIEM solution by Supplier 2

SIEM **Supplier 2 Product Y**	**Requirement**	**Weight** **(W)**	**Value** **(V)**	**Score** **(S)**	**Total**
PLATFORM	a_1	1.0	1	1	31.4
	a_2	1.0	4	4	
	a_3	0.9	1	0.9	
	a_4	0.8	1	0.8	
	a_5	1.0	4	4	
	a_6	1.0	0	0	
	a_7	1.0	2	2	
	a_8	1.0	4	4	
	a_9	1.0	1	1	
	a_{10}	1.0	4	4	
	a_{11}	1.0	0	0	
	a_{12}	0.8	2	1.6	
	a_{13}	0.8	1	0.8	
	a_{14}	0.8	1	0.8	

	a_{15}	1.0	1	1	
	a_{16}	0.8	2	1.6	
	a_{17}	0.8	1	0.8	
	a_{18}	1.0	0	0	
	a_{19}	1.0	0	0	
	a_{20}	0.9	2	1.8	
	a_{21}	0.6	1	0.6	
	a_{22}	0.7	1	0.7	
	a_{23}	0.5	0	0	
OPERATIONS	b_1	1.0	1	1	2.9
	b_2	1.0	1	1	
	b_3	0.9	1	0.9	
	b_4	0.8	0	0	
INTERGRATIONS	c_1	0.8.	1	0.8	1.6
	c_2	0.5	0	0	
	c_3	0.8	0	0	
	c_4	0.8	1	0.8	
	c_5	0.5	0	0	
ADVANCED FEATURES	d_1	0.7	0	0	0
	d_2	0.6	0	0	
	d_3	0.6	0	0	

	d_4	0.5	0	0	
LICENSING AND SUPPORT	e_1	1.0	2	2	9.6
	e_2	1.0	2	2	
	e_3	1.0	1	1	
	e_4	0.8	2	1.6	
	e_5	1.0	2	2	
	e_6	1.0	1	1	
				TOTAL SCORE	45.5

SUPPLIER 3, PRODUCT Z:

This solution is divided on three layers: collection, archiving and correlation.

(a) Collection layer: composed of software instances (agents) where each agent works on a source.

(b) Management layer: composed of hardware appliances to manage logs.

(c) Correlation layer: a copy of the collected logs is sent to this layer to be processed; it consists of one hardware and is able to handle fair number of EPS (real time correlation).

The solution ensures high availability through deploying redundant appliances. Other main requirements that this solution covers are indexing, hashing and normalization. However the license is activated at the beginning of the deployment process and there is no delayed license.

Table 9. Evaluation of the SIEM solution by Supplier 3

SIEM Supplier 3 Product Z	Requirement	Weight (W)	Value (V)	Score(s)	Total
PLATFORM	a_1	1.0	4	4	47.3
	a_2	1.0	2	2	
	a_3	0.9	2	1.8	
	a_4	0.8	2	1.6	
	a_5	1.0	4	4	
	a_6	1.0	4	4	
	a_7	1.0	4	4	
	a_8	1.0	2	2	
	a_9	1.0	1	1	
	a_{10}	1.0	4	4	
	a_{11}	1.0	2	2	
	a_{12}	0.8	2	1.6	
	a_{13}	0.8	1	0.8	
	a_{14}	0.8	1	0.8	
	a_{15}	1.0	1	1	
	a_{16}	0.8	2	1.6	
	a_{17}	0.8	1	0.8	
	a_{18}	1.0	2	2	
	a_{19}	1.0	4	4	
	a_{20}	0.9	2	1.8	
	a_{21}	0.6	1	0.6	
	a_{22}	0.7	2	1.4	
	a_{23}	0.5	1	0.5	
OPERATIONS	b_1	1.0	2	2	4.8
	b_2	1.0	1	1	
	b_3	0.9	2	1.8	
	b_4	0.8	0	0	
INTERGRATIONS	c_1	0.8	1	0.8	6.3
	c_2	0.5	1	0.5	

	c_3	0.8	1	0.8	
	c_4	0.8	4	3.2	
	c_5	0.5	2	1	
ADVANCED FEATURES	d_1	0.7	2	1.4	4.2
	d_2	0.6	1	0.6	
	d_3	0.6	2	1.2	
	d_4	0.5	2	1	
LICENSING AND SUPPORT	e_1	1.0	1	1	8
	e_2	1.0	2	2	
	e_3	1.0	2	2	
	e_4	0.8	0	0	
	e_5	1.0	2	2	
	e_6	1.0	1	1	
				TOTAL SCORE	70.6

SUPPLIER 4, PRODUCT T:

The solution's architecture is distributed in its operation, it includes:

(a) Collection layer composed of nodes for event collection, parsing, and filtering.

(b) Storage layer: composed of nodes for indexing and storing in database, and for rule/report computation.

(c) Processing and correlation layer: for processing tasks.

The solution has a scale out architecture for analytics, where the layers work in a co-operative fashion to provide scalability. A lot of the main requirements are compliant, however, there is no mention for normalization and hashing, the anonymization of logs is not supported, and the license for testing period is too short.

Table 10. Evaluation of the SIEM solution by Supplier 4

SIEM Supplier 4 Product T	Requirement	Weight (W)	Value (V)	Score (S)	Total
PLATFORM	a_1	1	2	2	36.7
	a_2	1	1	1	
	a_3	0.9	1	0.9	
	a_4	0.8	1	0.8	
	a_5	1	1	1	
	a_6	1	4	4	
	a_7	1	2	2	
	a_8	1	1	1	
	a_9	1	1	1	
	a_{10}	1	4	4	
	a_{11}	1	4	4	
	a_{12}	0.8	2	1.6	
	a_{13}	0.8	1	0.8	
	a_{14}	0.8	1	0.8	
	a_{15}	1	1	1	
	a_{16}	0.8	2	1.6	
	a_{17}	0.8	1	0.8	
	a_{18}	1	2	2	
	a_{19}	1	4	4	
	a_{20}	0.9	2	1.8	
	a_{21}	0.6	1	0.6	
	a_{22}	0.7	0	0	

	a_{23}	0.5	0	0	
OPERATIONS	b_1	1	1	1	5.4
	b_2	1	1	1	
	b_3	0.9	2	1.8	
	b_4	0.8	2	1.6	
INTERGRATIONS	c_1	0.8	1	0.8	7.1
	c_2	0.5	1	0.5	
	c_3	0.8	2	1.6	
	c_4	0.8	4	3.2	
	c_5	0.5	2	1	
ADVANCED FEATURES	d_1	0.7	2	1.4	4.2
	d_2	0.6	1	0.6	
	d_3	0.6	2	1.2	
	d_4	0.5	2	1	
LICENSING AND SUPPORT	e_1	1	2	2	7.6
	e_2	1	1	1	
	e_3	1	1	1	
	e_4	0.8	2	1.6	
	e_5	1	1	1	
	e_6	1	1	1	
				TOTAL SCORE	61

Therefore, applying the quantitative evaluation on the received SIEMs will output an evaluation (total score) for each solution. At the end, two solutions are selected as qualified SIEM solutions, where only one of them will be adopted finally. In the next section, *applicability* evaluation is applied to select one of the two qualified SIEM solutions.

6.4.2.2 Applicability evaluation: Using a qualitative method

As said before, not always the solution that got the highest score in the requirement-based evaluation will be adopted; however, as explained before, there are other factors and indicators that should be considered, and they are represented in the *applicability* evaluation. After evaluating quantitatively, the SIEM solution, the next step is to apply the qualitative evaluation (*applicability* evaluation). In our case, the applicability evaluation is applied to choose between the two qualified solutions (by supplier 1 and supplier 3). Tables 11 and 12 shows the applied qualitative evaluation:

SUPPLIER 1 PRODUCT X

Table 11. Qualitative evaluation for the first SIEM solution by Supplier 1

Supplier 1 Product X	Indicators	Importance or weight	Evaluation	NOTES
Applicability	***Level of Compliance***	High	Excellent	High compliance
	Complexity	High	Very Good	Very Flexible
	Quality of services/ capabilities	High	Very Good	
	Robust Architecture	High	Very Good	High availability is guaranteed
	Scalability	Medium	Good	
	Clearness/ complete description/ vision	High	Excellent	

Licensing and support services	***Installation duration/ clearness of road map***	High	Very Good	Clear roadmap including all steps
	Licensing	High	Very Good	Clear dimensioning including user acceptance test period
	Support	Medium	Good	24x7
	Training	Medium	Good	
Advanced Features	***Additional features***	Low	Good	
	Integration with third parties	Medium	Good	
Other indicators	***Expertise/Skill of Vendor/ Supplier***	Medium	Excellent	One major player
	Price	Medium	Excellent	
OVERALL RESULT				ACCEPTED

SUPPLIER 3, PRODUCT Z

Table 12. Qualitative evaluation for the third SIEM solution by Supplier 3

Supplier 3 Product Z	Indicators	Importance or weight	Evaluation	NOTES
Applicability	*Level of Compliance*	High	Very Good	Mostly compliant
	Complexity	High	Good	Non trivial deployment
	Quality of services/ capabilities	High	Very Good	
	Robust Architecture	High	Very Good	High redundancy is guaranteed
	Scalability	Medium	Good	
	Clearness/ complete description	High	Excellent	
Licensing and support services	*Installation duration/ clearness of road map*	High	Very Good	
	Licensing	High	Insufficient	Uses acceptance-testing period was not included and low dimensioning of EPS
	Support	Medium	Good	
	Training	Medium	Good	
Advanced Features	*Additional features*	Low	Good	
	Integration with third parties	Medium	Good	
Other indicators	*Expertise/Skill of Vendor/ Supplier*	Medium	Very Good	Major player but adopting less know system integrator
	Price	Medium	Insufficient	A bit high if compared to proposed number of licenses
OVERALL RESULT				REJECTED

Both the two SIEM solutions (by supplier 1 and 3) got the highest value. However, the first solution matched more the customer's needs and the defined indicators. Even if the solution proposed by supplier 3 got a higher value, it was not selected. The first solution offered two architectures to choose from, which helps in the ease of deployment, better licensing, support, and it matched another relevant indicator, which is the price, offering a solution that is consistent to the budget.

6.5 Comparison

To prove the eligibility of using our approach we select Gartner's work to compare with. Gartner Inc. is one of the leading information technology research and advisory company that deliver the technology-related insight necessary for clients to make the right decisions, research, analyze and interpret the business of IT within the context of their individual role.

Gartner published its 2018 Critical Capabilities for Security Information and Event Management report [35] to evaluate and rank different SIEM vendors (products). First, they started by specifying the critical capabilities that a SIEM should have. Critical capabilities are architecture, deployment, operations and support, log and data management, real-time monitoring, analytics, data and application monitoring, threat and environmental context, user context and monitoring, incident management and threat detection tools.

After that, they defined three use cases: basic security monitoring, complex security monitoring, and advanced threat defense, where the SIEM vendors are evaluated in each use case. These use cases used represent three "operational" levels in which the SIEM will be deployed.

Critical capabilities are then weighted, where each capability is given a weight in terms of its relative importance in each use case. In addition, for each SIEM product, the critical

capabilities are rated on a scale of 1 to 5, where 1 is a poor requirement, whereas 5 is an outstanding (significantly exceeds requirements). Finally, an overall score is carried out for each SIEM on the three different use cases. To determine an overall score for each product in the use cases, the ratings (evaluations) are multiplied by the weightings to come up with the product score in use cases. Different SIEM products developed by the most expertise and skilled vendors are used in the evaluation applied by [35], where each vendor's product or service is evaluated in terms of how well it delivers each of the capabilities defined. This kind of evaluation can be defined as a product driven evaluation, where the capabilities of the product are evaluated without differentiating the customer's needs factor.

On the other hand, our approach suggests the customer to create a more detailing list of requirements that reflects their needs, which is a subjective and detailed list representing the SIEM seeking for, therefore, to receive an overall solution and not just a product, and so the evaluation will be influenced by these specific requirements.

Moreover, we defined a second level of evaluation, which is a qualitative evaluation method (applicability based) that selects from the qualified solution evaluated previously by the quantitative evaluation method (requirement based).

Table 13. A comparison between the proposed SIEM evaluation approach and Gartner's report

Product-driven SIEM evaluation	**Customer-driven: Our SIEM Evaluation Approach**
Suggests a list of the main requirements (capabilities) that a powerful SIEM should have	Suggests a list of requirements with some specifications from a customer point of view
Has a flat approach when weighting the requirements	Can adopt weights to reflect more significant requirements
Uses quantitative evaluation	Uses both quantitative and qualitative evaluation

Evaluation is applied for general use cases	Evaluation is applied based on customer needs and context
Evaluates a product	Evaluate an overall solution: product, architecture, deployment, price, etc…

In conclusion, organizations tend to ensure good management controls are in place to avoid unexpected outcomes and to keep on track, so they require a structured approach for managing their tasks. This chapter proposed a thorough approach to support companies that are seeking to adopt SIEM systems into their environments; it suggests suitable technological and business requirements that are believed to be valuable in a SIEM system and proposes a two-phase evaluation process to measure the compliance and applicability of a SIEM. At the end, as said before, this approach must be completed by a testing -phase of the selected SIEM to confirm that the received requirements are as described by the suppliers.

7 Chapter seven: Risk Analysis Using PILAR software: An instructional document to use PILAR

Risk analysis is one of the CSAC department's roles at Hitachi rail STS Company. Risk analysis is the process for estimating the risks to which the system's assets are exposed to, and then should be followed by the treatment phase. In the field of cyber security, there are a lot of methods used, and in this chapter we aim to find a methodical way to conduct a risk analysis study which takes into account cyber and physical sides of the system under study. This part of the work is represented as an instructional document that is also a part of the ISMS and can be used in the future referring to it in case there is need to know how risk analysis is applied at the company.

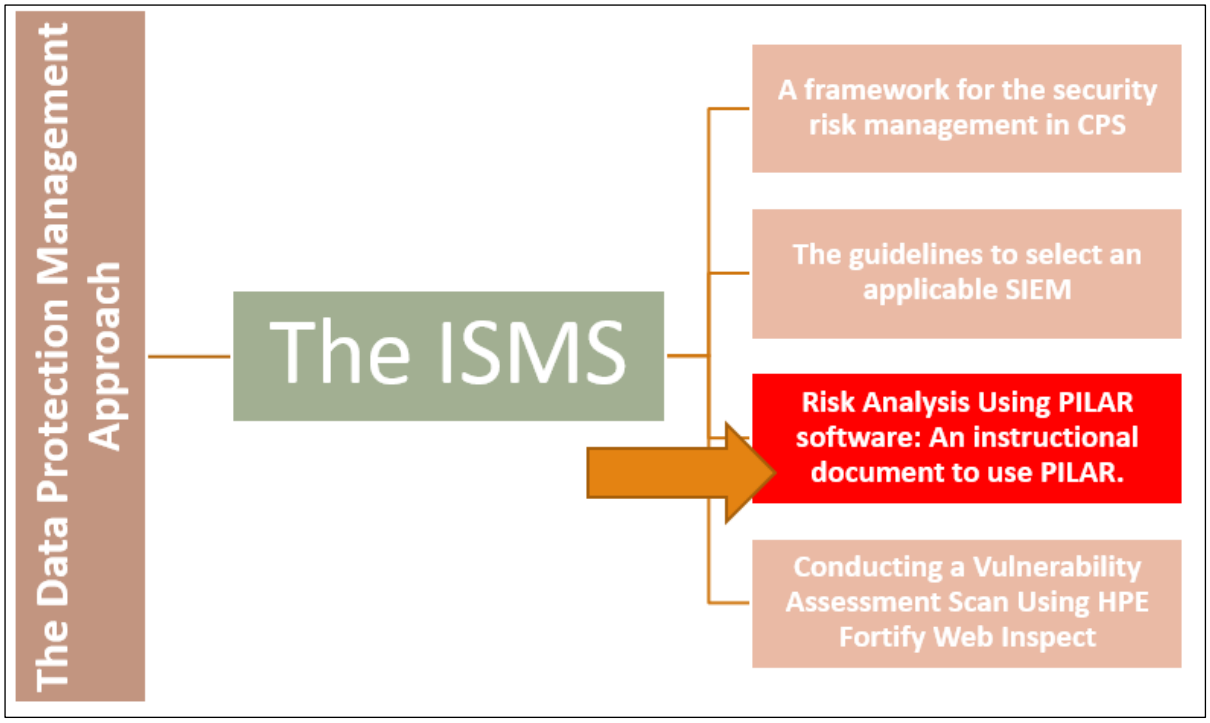

Figure 26. PhD Goals (related to the ISMS): Risk Analysis Using PILAR

There are several approaches for the problem of analyzing the risks: informal handbooks, methodical approaches or supporting tools, where all provide a guide for risk analysis. However, methods might differ in some steps, or in the way of identifying and valuating the assets or threats. Some are basically used in cyber security of information systems, and others

can be used in physical security. The great challenge of these approaches is the complexity of the problem they have to face; complexity in the sense that there are many elements to be considered and, if it is not done rigorously, the conclusions will be unreliable.

7.1.1 MAGERIT Methodology: Objectives and steps

One of the methodical approaches is MAGERIT method suggested by Enisa [27], which is the European Union Agency for Network and Information Security. MAGERIT is an open methodology for Risk Analysis and Management, developed by the Spanish Ministry of Public Administrations.

Magerit is a risk analysis and management methodology for information systems developed by CSAE (Consejo Superior de Administraci´on Electr´onica), Spain [28]. It was published in 1997. Magerit v2 was published in 2005, and Magerit v3 was published in July, 2014**.**

Vendor name: Ministerio de Administraciones Publicas (Spanish Ministry for Public Administrations).

Following a methodical way in a risk management study is important in order to obtain an efficient study. The objective of Magerit method is to cover both risk analysis and risk treatment for a thorough risk management. The ultimate aim of using MAGERIT is to make a methodical approach that leaves no room for improvisation, and not to depend on the analyst's whim.

The purpose of Magerit is directly related to the generalized use of IT systems, communications, and electronic media, which bring evident benefits for the users but which is also subject to certain risks that must be kept under control by means of security countermeasures that generate confidence in the use of these media. This method follows the international concepts as in:

1. ISO 31000:2009 – Risk management – Principles and guidelines.

   According to ISO 31000 terminology, Magerit responds to what is called "Risk Management Process", the section "Implementing Risk Management" within the "Framework for Risk Management".

2. ISO/IEC 27005:2011 - Information technology - Security techniques - Information security risk management.

Next section lists the steps of Magerit methodology.

7.1.2 MAGERIT method steps

MAGERIT offers a systematic method for analyzing risks, and helps in describing and planning the appropriate measures for keeping the risks under control. And finally, prepare the organization for the processes of evaluating, auditing, certifying or accrediting, as relevant in each case.

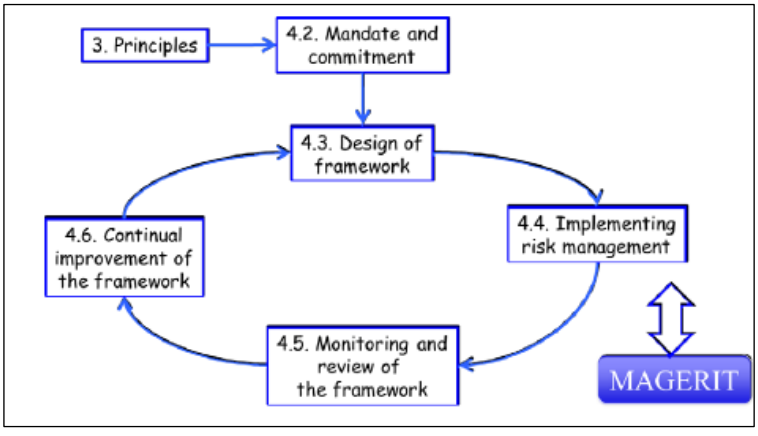

Figure 27. ISO 3100- Framework for risk management

It also aims to achieve uniformity in the REPORTS containing the findings and conclusions from a risk analysis and management project through the following steps:

1. Value mode (Identification and valuation of assets): First step is to identify the assets composing the system or the organization and characterizes them according to their type. And then describe the value of the assets (critical, very important, important…) for the organization as well as the dependencies between the various assets.
2. Risk map: Identify the threats to which the assets are exposed, and evaluate them. This activity identifies the relevant threats on the system and characterizes them according to the estimates of occurrence (likelihood).
3. Statement of applicability (define available safeguards): For a set of safeguards indicate which ones are applicable in the information system under study, and which ones are meaningless. Given the wide variety of safeguards, it is necessary to go through all of them and retain (keep) only those which are relevant for protecting the asset.
4. Safeguard evaluation: Evaluation of the effectiveness of the existing safeguards in relation to the risks systems face. In other words, check the effectiveness of the existing safeguards deployed in the system and how they would behave against the threats we define (report on the safeguards deployed, characterized by their degree of effectiveness).
5. Risk status: Classification of the assets by their residual risk; that is to determine the residual impact on the system, by what could happen, taking the safeguards used into consideration.
6. Output: Report detailing for each asset the impact and the potential and residual risks regarding every threat.
7. Deficiencies report (Vulnerabilities report): Absence or weakness of the safeguards that appear as appropriate to reduce the risks to the system. A report identifying the

deficiencies and vulnerabilities in the system. (e.g.: what protection should be available and it is not currently).

8. Compliance (Meeting some requirements): Formal statement that it is in line and in accordance with the corresponding regulations. Some organization has its own established regulations which must be taken into account.
9. Security plan: Group of security programs (master plans or strategic plans) that put the risk treatment decisions into action. Decision-making concerning risk treatment may lead to recommend safeguards assessing its influence on impact and risk indicators. Security plan when implemented and operated, must meets the proposed objectives with the level of risk accepted by the Management.

7.1.3 PILAR software

PILAR software is used to perform the steps of the MAGERIT methodology. It's GUI (graphical user interface) enables the user to execute the MAGERIT in an understandable and easy way. The tool provides fast calculations and generates a quantity of textual and graphical reports [29]. Note that running PILAR requires a license and a JDK to be installed on the PC.

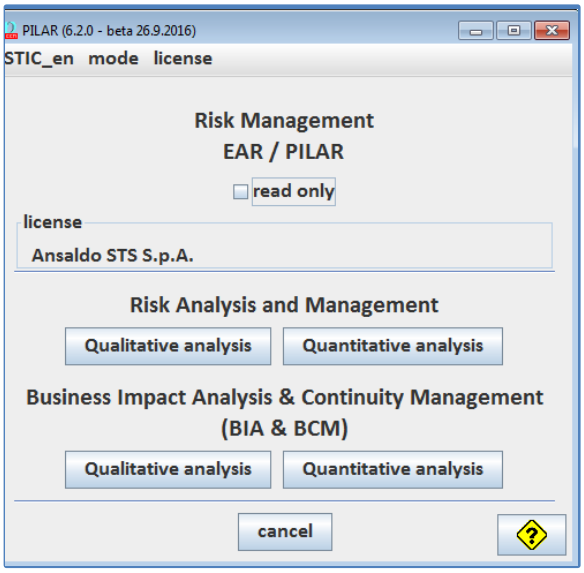

Figure 28. PILAR: First Screen

PILAR software has been funded by the Spanish National Security Agency. It is designed to support the risk management process along long periods, providing incremental analysis as the safeguards improve. Its functionalities include mainly:

- Quantitative and qualitative Risk Analysis and Management.
- Quantitative and qualitative Business Impact Analysis & Continuity of Operations.

PILAR enables the user to create a project, identify the assets for the system under study, generate threats and safeguards and other functionalities. The figure below is for creating a new project:

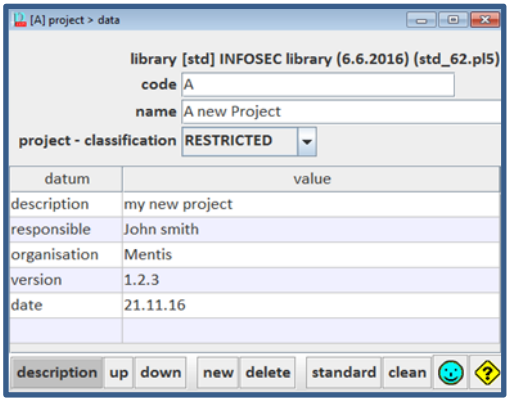

Figure 29. PILAR: Create New Project

Creating a new project requires at least a code and a name. Other details could be added such as a description, responsible person, organization version and date.

The user can change from the list to choose the level and the options that will be presented.

Basic	Only basic options, with the aim of simplifying life to early users.
Medium	Somewhere in between basic and expert
Expert	All the options are shown

Figure 30. PILAR: Level of user

The user can also modify the behavior of PILAR in several aspects from the Edit>Options menu.

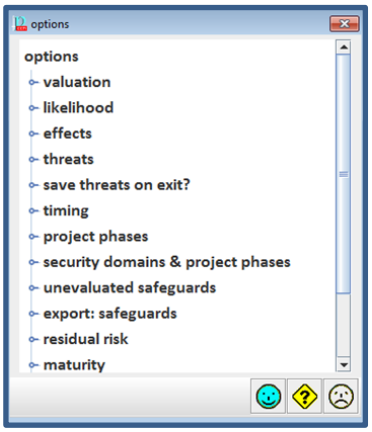

Figure 31. PILAR options

These options are specific for each project analysis, so editing options is done only when a project is open, and the options will affect only the current project. Some personalized versions of the tool may offer additional options. For example, the valuation of assets may include dependencies or domains, so the system may be rated asset by asset (plus dependencies) or by security domains.

On the other hand, the identification of the threats can be manual, automatic or mix. And other options could be changed like the likelihood type (frequency, level, potential…)

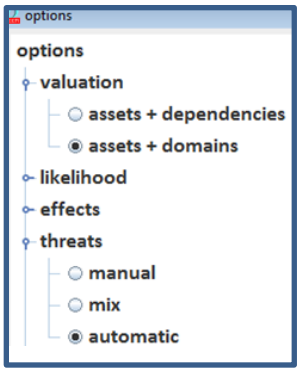

Figure 32. PILAR: Change options

The department of Cybersecurity Assurance and Control at Hitachi Rail STS Company uses PILAR software to conduct risk analysis studies. For this goal, a document was created that includes all the instructions that should be followed in any such study by the department.

The work was demonstrated by writing a 70-pages document about using PILAR in any risk assessment project in the future. This document is used as an internal instructional document of the company; it will be a part of the Information security management system (ISMS) of the company.

8 Chapter eight: Conducting vulnerability assessment scans Using Fortify Web Inspect

Vulnerability assessment is also one of the CSAC department's roles at Hitachi rail STS Company. Vulnerability assessment is the process of defining, identifying, classifying and prioritizing vulnerabilities in computer systems, applications and network infrastructures and providing the organization doing the assessment with the necessary knowledge, awareness and risk background to understand the threats to its environment and react appropriately. Vulnerability assessments depend on discovering different types of system or network vulnerabilities, which means the assessment process includes using a variety of tools, scanners and methodologies to identify vulnerabilities, threats and risks.

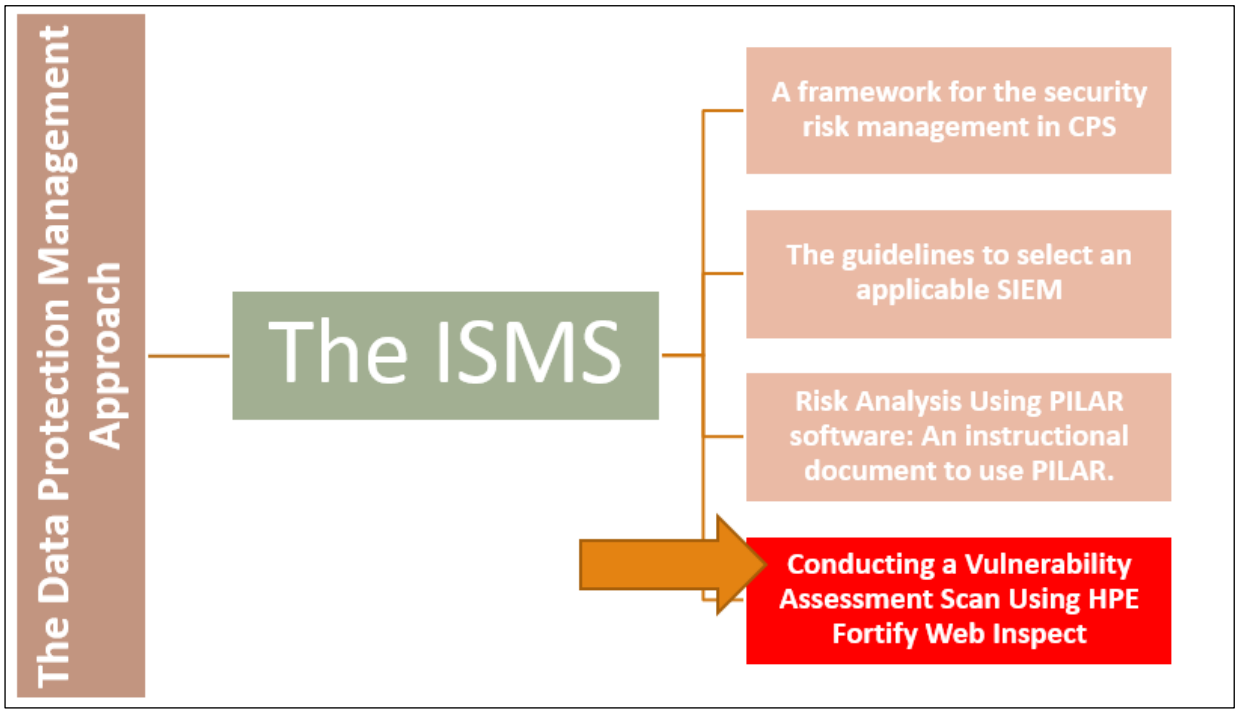

Figure 33. PhD Goals (related to the ISMS): Conducting a vulnerability assessment using Fortify WebInspect

Some of the different types of vulnerability assessment scans include the following:

- Network-based scans are used to identify possible network security attacks. This type of scan can also detect vulnerable systems on wired or wireless networks.
- Host-based scans are used to locate and identify vulnerabilities in servers, workstations or other network hosts. This type of scan usually examines ports and services that may also be visible to network-based scans, but it offers greater visibility into the configuration settings and patch history of scanned systems.
- Wireless network scans of an organization's Wi-Fi networks usually focus on points of attack in the wireless network infrastructure. In addition to identifying rogue access points, a wireless network scan can also validate that a company's network is securely configured.
- Application scans can be used to test websites in order to detect known software vulnerabilities and erroneous configurations in network or web applications.
- Database scans can be used to identify the weak points in a database so as to prevent malicious attacks, such as SQL injection attacks.

In this chapter, we aim to show how vulnerability assessment could be applied for web applications using a vulnerability scanner. Web Application Vulnerability Scanners are automated tools that scan web applications, normally from the outside, to look for security vulnerabilities such as Cross-site scripting, SQL Injection, Command Injection, Path Traversal and insecure server configuration.

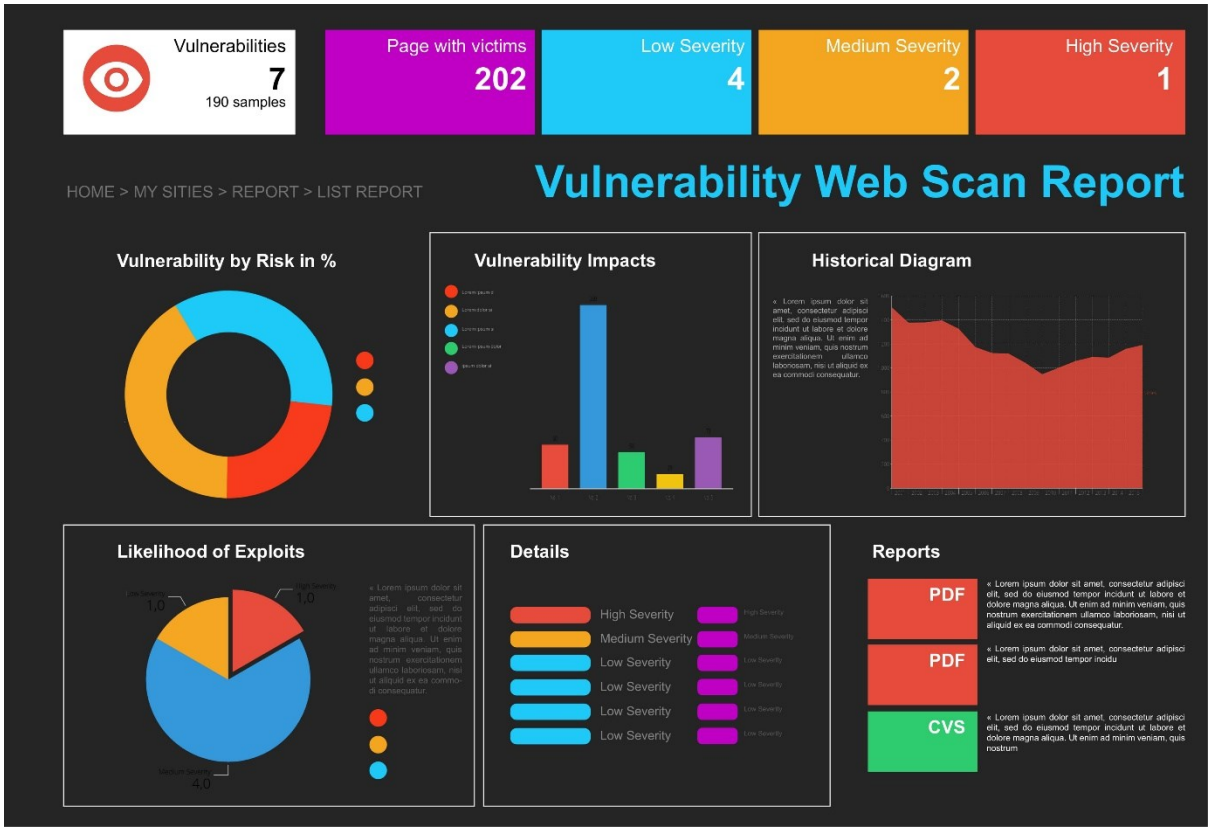

Figure 34. An example for the dashboard for a web application vulnerability scanner

This part of the work defines one of the web application vulnerability scanners, called “Fortify web inspect”, this part of the work was concluded by an instructional document that was also a part of the ISMS and can be used in the future. HPE Security Fortify WebInspect is an automated Web application and Web services vulnerability scanning solution. As you initiate a scan, Fortify WebInspect assigns agents that dynamically catalog all areas of a Web application. These agents report their findings to a main security engine that analyzes the results. Fortify WebInspect then launches "Threat Agents" to evaluate the gathered information and apply attack algorithms to determine the existence and relative severity of vulnerabilities. With this smart approach, Fortify WebInspect continuously applies appropriate scan resources that adapt to your specific application environment.

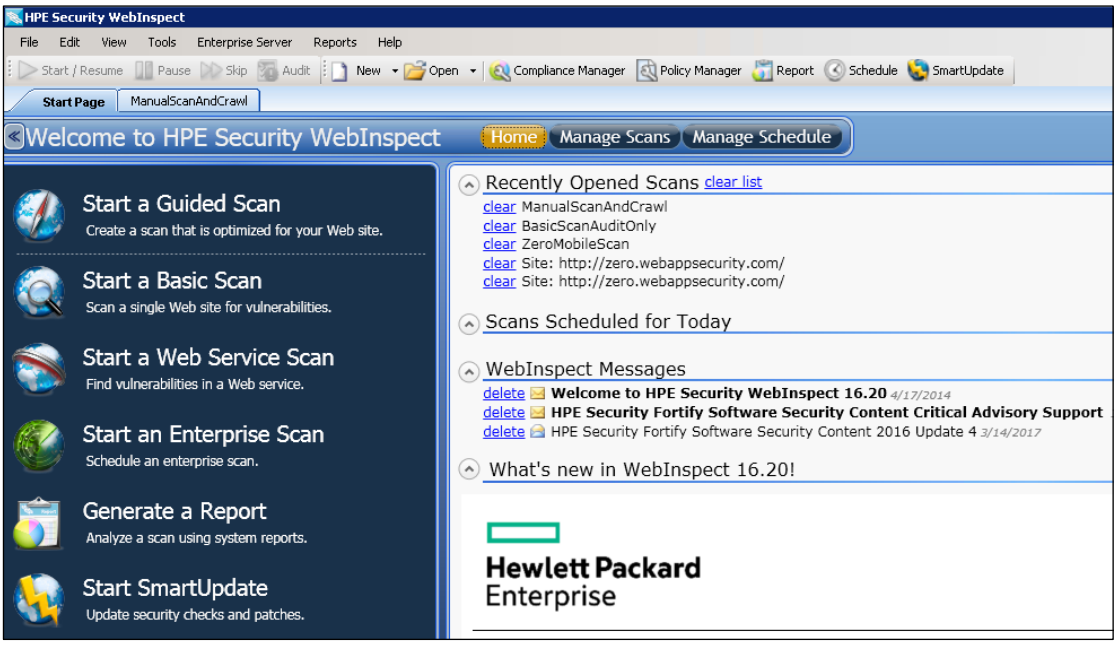

Figure 35. Fortify Web Inspect Start Page

Fortify WebInspect employs an extensive arsenal of attack agents designed to detect security flaws in web-based applications. It probes the system with thousands of HTTP requests and evaluates each individual response. This session-based assessment reports each vulnerability, pinpoints its location in the application, and recommends corrective actions that should be taken. It is, basically, a quantitative analysis of the system.

Fortify WebInspect has a set auditing policies, settings, and other configurations that changes the behavior and its way of functionality. So there is a need to follow certain set of settings and policies that best fits scanning a target URL, website or a webservice.

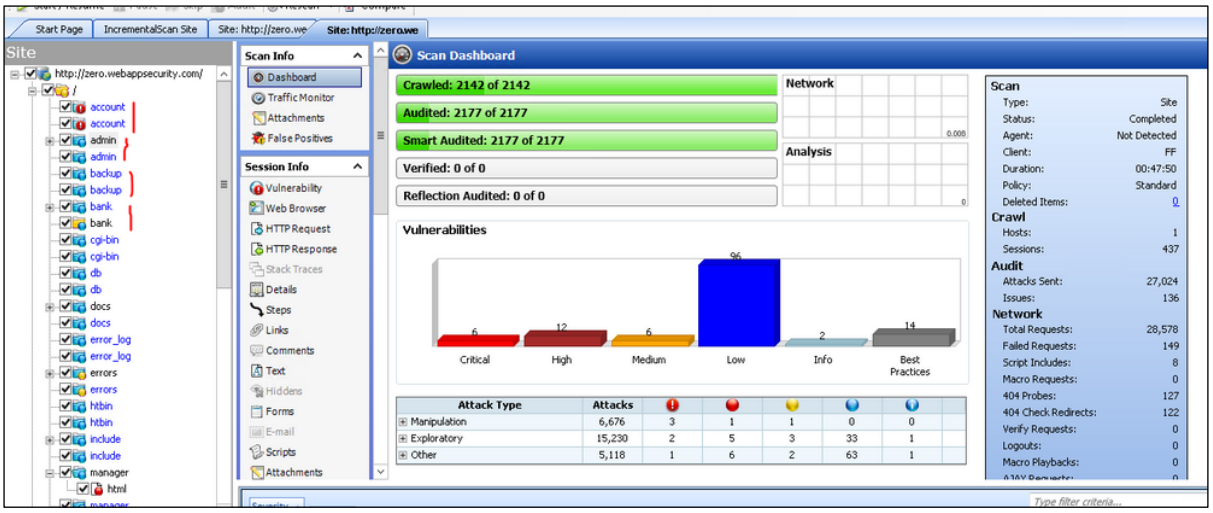

Figure 36. Fortify Dashboard after conducting a scan

The work was demonstrated by creating an instructional document about using this software in the vulnerability assessment projects. This document will be used as an internal document for the company; it will be also a part of the Information security management system (ISMS) of the company.

9 Chapter nine: Classifying web attacks using network statistical features and machine learning: Analyzing the behavior of the attack generator

Web attacks arise and continue evolving day after day; they could be of different types and targeting diverse assets. In order to ensure the security against web attacks, many procedures, appliances and countermeasures might be developed and installed. Firewalls, intrusion detection and prevention systems are security appliances that could be used to detect them. Such appliances might be signatures based detection, or even anomaly based detection that use machine learning techniques to analyze the network. Both detection of the anomalies and their type, could be accomplished by using machine-learning techniques applied on a relevant dataset. A lot of work was done in the field of intrusion detection and classification of web attacks, however, few if no one addressed the detection of the tool type used by the attacker. Detecting the type of the tool used to generate a web attack could be worthy and might aid in the prevention and mitigation of web attacks, since any information about the attackers might be helpful and worth a lot to the cyber security analysts. In this work, our goal is to analyze the behavior of the attacker; by not only detecting the type of the attack, but also creating a second classification layer to detect the type of the software used by the attackers. This chapter presents a method to analyze the behavior of the attacker by means of classifier models. The principle of operation is to deploy conventional classifier techniques for discriminating the categories of attacks, and, at the same time, to implement an additional classification layer to detect the type of software used by the attacker. The chapter shows empirically that it is possible to discriminate different tool by using computationally light, white-box models. As a result, analysts and network managers can better understand the determinist rules extracted from data and use them to enhance security policies and systems.

The use of computers, storage, networking and other devices to create, store and exchange all forms of data has increased significantly in the last years. Interconnectivity and data generated by devices has resulted in ‘an unprecedented improvement in the quality of life’ [44]. Particularly, the use of websites and web applications has also increased in recent years especially by companies and critical infrastructure applications that plays a key role in economy and society. The number of websites has reached an unprecedented number in the last few years since the World Wide Web project was born in 1989. Both websites and web applications have become a basic part of business in nowadays.

Websites are collections of web pages, they can be accessed using a browser to read and look at everything that's on a website, they provide and present content for readers. On the other hand, web applications are defined by being interactive. A web application is used in order to perform a function and use some specific features, it can be defined as a web site that does more than displaying content, it has a business logic**.** It’s intended for user interactions*,* performing actual business functions**.** Using websites and web applications, businesses can now develop and reach its goals much faster. Moreover, businesses might obtain an effective economic growth if they adopt an appropriate one. Good Websites and web applications can play a crucial role in the branding process. Using them, the opportunity of selling the services or products is increased; they represent the first communication channel between potential customers and the business organization, and at the same time, popularity of the organization gets boosted and lead generation improves. Small businesses, banks, and major industrial concerns all depend on web applications. Websites are the public face of both business and government, so any issues with a website can damage reputation. For this and other reasons, web application owners are motivated to improve and support their sites. Proper attention to

cybersecurity is a key part of any such strategy. Companies, organizations, critical infrastructure, governments, or even small businesses are all using web applications to manage an importance part of their business. Nowadays, there are over 1.5 billion websites on the World Wide Web today. Of these, less than 200 million are active. By "**Website**", we mean unique hostname (a name which can be resolved, using a name server, into an IP Address) [45] [46].

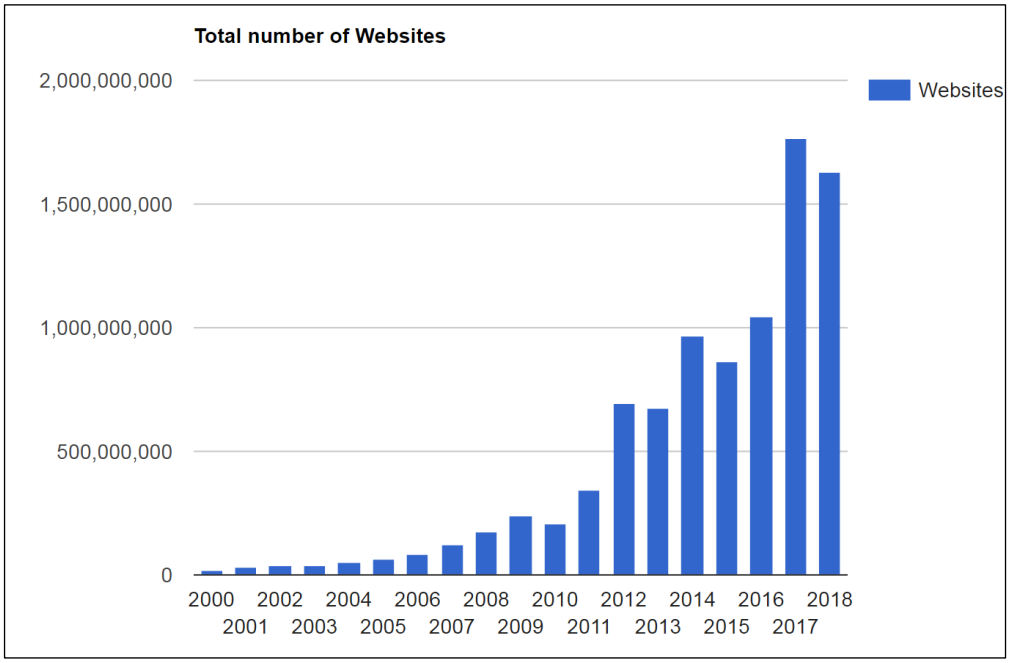

Figure 37. Total number of websites. Source: NetCraft and Internet Live stats

In parallel, web attacks are evolving and targeting vulnerable and sensitive sites, which may cause unacceptable threats especially for systems that are part of the critical infrastructure or have a high value business. In most of the cases, websites are usually connected to databases that might contain confidential information, if attackers scanning sites for vulnerabilities gain unauthorized access to the sites, it might lead to hazardous situations with catastrophic consequences.

In addition, according to SiteLock, a global leader in website security [47], approximately 6% of websites—up to 113 MILLION WEBSITES globally—have a security vulnerability. In Q1

2018, SiteLock surveyed over 250 website owners to assess their knowledge and what they fear most about website security. Almost half (46%) of website owners surveyed reported that their website was the victim of a security incident in 2017. These responses serve as a reminder that a website security attack can happen at any time to anyone. Of those that did report a cyberattack on their website, 36% reported that the incident caused lost revenue and harmed their bottom line. A staggering 42% of respondents also reported that their biggest website security fear was a defacement, indicating a lack of awareness that the quieter and stealthier malware attacks are just as, if not more, damaging [48].

A lot of strategies might be adopted to secure websites and web applications against web attacks, one way is intrusion detection using machine learning techniques, where a network intrusion detection system is used aiming to classify the monitored traffic as either "legitimate" or "malicious, and sometimes to detect the type of the malicious traffic. Both detection of the anomalies and their type could be accomplished by using machine learning applied on an appropriate dataset. Intrusion detection by using machine-learning techniques is a classical solution to secure websites and web applications. Intrusion-detection systems classify the observed traffic as either legitimate or malicious, and sometimes can detect the type of the malicious traffic. These goals could both be accomplished by using machine-learning tools, trained with a suitable dataset. Many datasets such as DARPA98, KDD99, ISC2012, and ADFA, exist for that purpose; these datasets contain data corresponding to normal and malicious traffic, and mostly support either intrusion detection or to attack-type recognition. Despite their effectiveness, machine-learning methods have two major drawbacks: first, accurate models typically require a considerable amount of memory and floating point operations. Secondly, in most cases human users cannot interpret the actual classification principles applied by the models (an issue known as a black-box phenomenon).

The research presented in this chapter aims to get precious side-information about the attackers that might help the manager in both the mitigation and the prevention process. This is attained by detecting the actual type of software used in staging web attacks.

The first step of this approach is to gather a significant dataset of web attacks for addressing this aspect; hence, the training set should cover not only the attack type but also info about the exploiting tool. To maximize coverage, that dataset should replicate/cover the most significant attacks against web applications; each attack scheme should be implemented by using different tools to encompass the possible behaviors of the attacks.

Then, one trains a two-layer classification model to reach the combined goal; the first layer detects the type of the attack, whereas the additional classifier classifies the attacking tool. The latter layer holds a classification tree; that straightforward model has been selected because of its limited computational cost; more importantly, human supervisors can inspect and interpret the overall decision process.

The major contribution of this chapter can be summarized as follow:

- A novel dataset for attacking tool classification containing attacks generated by the most common tool for attacks generation;
- An effective strategy for attacking tool classification;
- A strategy for online classification based on low cost/human interpretable models.

9.1 Background and Related work:

Analyzing the network traffic, allow network defenders to perform a variety of analysis and can help them understand more what is going on in their network. For example, when examining web traffic, network flow data would contain the source and destination IP addresses involved, the amount of data sent, the number of packets, and the time duration of the communication, and much more features. Analyzing the network traffic represents a

wealthy process that can be employed for the purpose of understanding how everything on the network works by examining the statistical and characteristic behavior. Both detection of the anomalies and their type could be accomplished by using machine learning. Many datasets exists to help in the detection and classification of anomalies, they include DARPA98, KDD99, ISC2012, and ADFA 13 that have been used by researchers to detect intrusions and their types. Some of these datasets suffer from lack of traffic diversity and volumes, some of them do not cover the variety of attacks, while others anonymized packet information and payload which cannot reflect the current trends, or they lack feature set and metadata [49]. On the other hand, other datasets cannot be shared due to privacy issues or others are out of date because of the continuous change in attack strategies. There exist a number of datasets used in the field of network security and network security. As our work deals with web attacks in particular; in the next two subsections we list the most emerging web attacks nowadays, and present the most used datasets in similar works.

9.1.1 Backgrounds: Web applications and web attacks

Among the most critical security risks affecting web applications nowadays, the Open Web Application Security Project, or OWASP, defines the top 10 security risks to web applications. OWASP is an international non-profit organization dedicated to web application security. One of their best-known projects is the OWASP Top 10. The OWASP Top 10 is a powerful awareness document for web application security. It represents a broad consensus about the most critical security risks to web applications. Project members include a variety of security experts from around the world who have shared their expertise to produce this list. The OWASP TOP 10 security risks are Injections and Broken Authentication, Broken Access

Control, Security Misconfiguration, Cross-Site Scripting (XSS), Insecure Deserialization, Using Components With Known Vulnerabilities, Insufficient Logging And Monitoring [50].

On the other hand, in a study in 2019 about cyber threats, “Positive Technologies” [51], a leading global provider of enterprise security solutions, stated that web application hacking is one of the most frequent attacks on both organizations and individuals. Hacked sites can be used for a multitude of things: distributing malware, stealing data, posting ads or forbidden information, committing fraud, or penetrating an internal network.

Moreover, according to the Positive Technologies 2019 report that lists the main web attacks reported in the year, brute force, sql injection, path traversal, local file information, leakage, XSS, and Denial of Service (DOS) were among the most common attacks affecting websites [52].

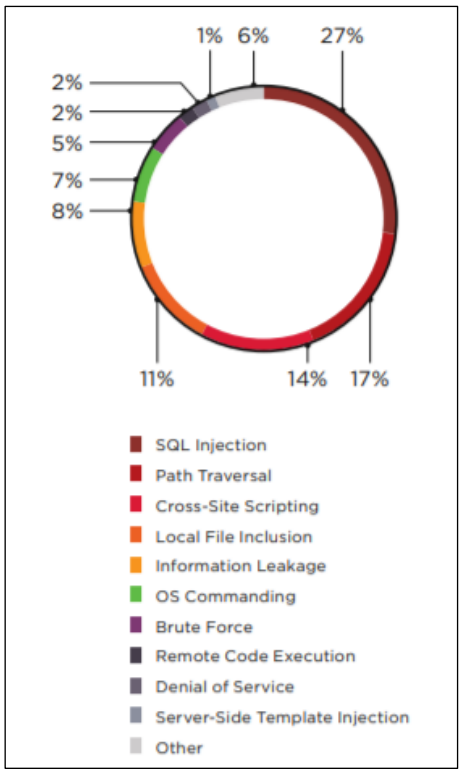

Figure 38. Top 10 web application attacks reported by “Positive Technologies”

9.1.2 Intrusion detection Using Machine Learning techniques //paraphrase

From the theoretical point of view, network intrusion detection systems mainly aim to classify the monitored traffic as either “legitimate” or “malicious.” Therefore, machine learning approaches are appropriate to solve such problems; and they have recently been widely applied to help better manage network intrusion detection issues. Machine learning (ML) is a field of artificial intelligence, which refers to a set of techniques that give computer systems the ability to “learn.” Typically, machine learning algorithms, such as artificial neural networks, learn from data samples to categorize or find patterns in the data, and enable computer systems to make predictions on new or unseen data instances based on the discovered patterns [53].

Machine Learning is a way of making a computer learn and take decisions without being explicitly programmed [54]. Machine learning techniques work by establishing a model that enables the analyzed patterns to be categorized. In general, machine learning techniques are able to deal with three common problems: classification, regression, and clustering. Network intrusion detection can be considered as a classification problem. Therefore, a labeled training dataset is usually required for system modeling. In Supervised Machine Learning [52], the learning algorithms tackle classification problems, and network intrusion detection relates to this context. A labeled training dataset is required for system modeling. The overall process includes three main steps, as shown in Fig.39. [55].

Conversely, unsupervised learning does not require a labeled sample, as the main purpose is to set up a description of the distribution of empirical data, to identify significant patterns or clusters of homogenous regions of the data space. In network security applications, unsupervised algorithms implement behavioral analysis and mostly aim at detecting zero-day attack scheme that are not covered by existing data sets.

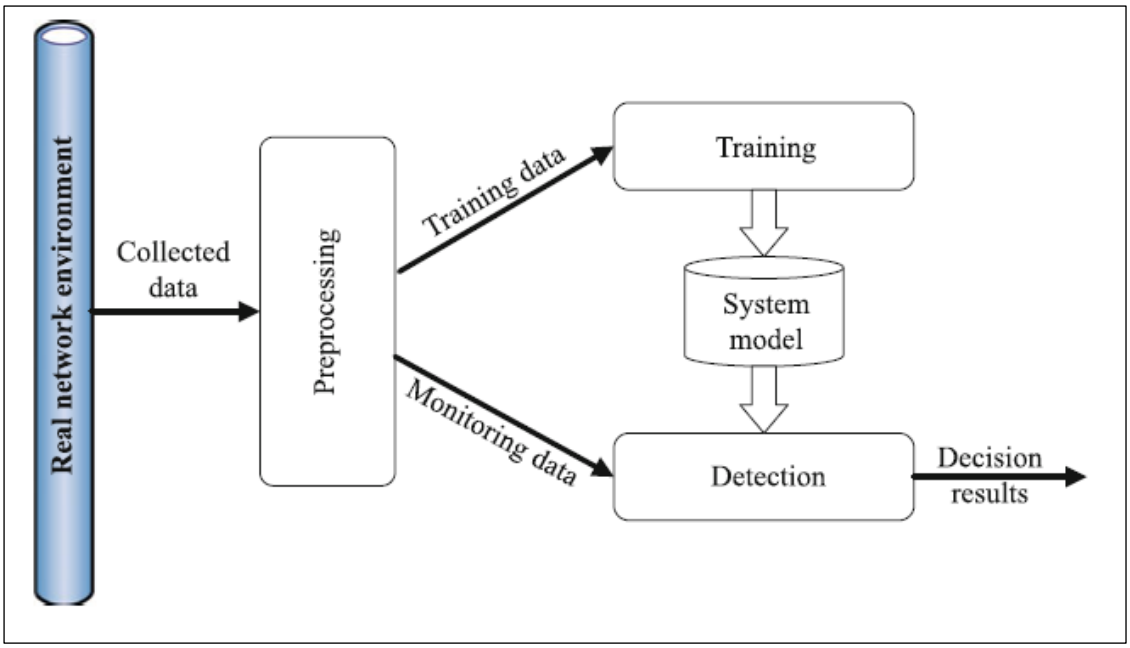

Figure 39. Intrusion detection process using machine leaning techniques

1. ***Preprocessing***: the data instances that are collected from the network environment are structured, which can then be directly fed into the machine learning algorithm. The processes of feature extraction and feature selection are also applied in this phase.
2. ***Training*****:** a machine learning algorithm is adopted to characterize the patterns of various types of data, and build a corresponding system model.
3. ***Detection*****:** once the system model is built, the monitored traffic data will be used as system input to be compared to the generated system model. If the pattern of the observation is matched with an existing threat, an alarm will be triggered.

Both supervised and unsupervised machine learning approaches have already been utilized to solve network intrusion detection problems. For instance, supervised learning-based classifiers have been successfully employed to detect unauthorized access, such as k-nearest neighbor (k-NN) [56], support vector machine (SVM) [57], decision tree [58], naïve Bayes network [54], and random forests [59]. In addition, unsupervised learning

algorithms, including k-means clustering [60] and self-organized maps (SOM) [61], have also been applied to deal with network intrusion detection problems, with good results.

9.1.3 The Available Datasets:

The two publicly available datasets that have provided something of a standardized setting in the past—the DARPA/Lincoln Labs packet traces [62], [63] and the KDD Cup dataset derived from them [64]—are now a decade old, and no longer adequate for any current study. The DARPA dataset contains multiple weeks of network activity from a simulated Air Force network, generated in 1998 and refined in 1999.

DARPA and KDD could be said to be the most used datasets in the network intrusion detection problems. DARPA (Lincoln Laboratory) was created in 1998-99 for network security, however it does not represent real-world network traffic, and contains irregularities. Also, it is outdate for the effective evaluation of IDSs on modern networks, both in terms of attack types and network infrastructure. Moreover, it lacks actual attack data records [65] [66]. On the other hand, The KDD'99 was used in the 3rd knowledge Discovery and Data Mining Tools Competition for building a network intrusion detector, is an updated version of the DARPA98. Thiss dataset has a large number of redundant records and is studded by data corruptions that led to skewed testing results [67]. Both DARPA and KDD have just 4 categories of attacks and lacks for some important ones such as SQL Injection, XSS, brute force attacks. Other datasets include CAIDA (Center of Applied Internet Data Analysis 2002-2016) that is not an effective benchmarking datasets due to a number of shortcomings, see [68] [69] [70] [71] for details. Also CDX (United States Military Academy 2009): suffers from the lack of traffic diversity and volume [72]. Moreover most of those datasets are old and they lack for web attacks especially SQL injection, cross-site scripting, and brute force. In [49], Lashkari et al. Generated

a new IDS dataset namely CICIDS2017 which covers common web attacks such as DOS, Brute force, SQL Injection, cross-site scripting. The dataset is completely labelled and more than 80 network traffic features extracted and calculated for all benign and intrusive flows by using CICFlowMeter software, which is publicly available in Canadian Institute for Cybersecurity website [73].

In conclusion, most of the available datasets lack relevant attack categories. Some of them group web attacks into one class and do not differentiate between the various schemes. When the dataset contains indeed a variety of web attack categories and is labeled accordingly, the samples do not cover any info about the software tools used to stage the attacks.

9.2 A classification layer to analyze the behavior of attacking tool:

Cyber security analysts try to collect any information about the attackers, to help in the mitigation of the risks. In particular, knowing the type of the software used in attacks would aid a lot. Since an attack could be generated in various ways, it is clear that the attack's behavior would be different based on the used tool. Even if same attack is applied, each attacking tool might have a specific behavior; each tool has a specific duration for the attack, specific number of packets per connection flow, protocol, or even the payload content, which might lead to a new different behavior on the network level.

Looking at simple statistical properties of protocol messages, such as statistics of packet inter-arrival times and of packets' size may be useful to perform monitoring actions [74]. "The key idea is that the information carried by packets at the network layer, such as packet-size an inter-arrival time between consecutive packets, are enough to infer the nature of the application protocol that generated those packets [75]." According to the last research and proposed evaluation framework by [76], eleven characteristics, namely Attack Diversity, Anonymity, Available Protocols, Complete Capture, Complete Interaction, Complete Network

Configuration, Complete Traffic, Feature Set, Heterogeneity, Labelling, and Metadata are critical for a comprehensive and valid Intrusion detection systems (IDS) dataset.

The first step in the proposed work of this chapter is to gather an appropriate dataset for a network traffic that represents the attacking scenario as whole, encompassing the statistical features characterizing both the attack type and the tool used. In this work, the training dataset holds web attacks that are believed to be most frequent against web applications. The sample includes various categories of web attacks, generated by using a set of different tools. This aims to get the widest coverage of possible behaviors of attacks, and therefore to aid in the analysis the behavior of the tools used to generate the attacks.

The research presented here focuses on web attacks for several reasons:

- The increased number of web attacks
- The diversity of tools available to demonstrate webattacks especially penetration testing tools
- The availability of online web applications and websites for penetration testing and vulnerability assessment.

In our dataset, a group of web attacks will be used to represent the main threats affecting web applications. Those web attacks are Brute Force, SQL Injection, XSS, and DOS. Those attacks are selected based on both the OWASP top 10 security risks and Positive Technologies 2019 report. In addition, these web attacks are believed to be enough to reach our goal in analyzing the behavior of the attack generator, in particular the tools used.

The procedure includes into 3 steps. The first step involves the simulation of Web Attacks, to create a dataset of web attacks containing the four main categories selected above. The simulation relies on a set of tools, whose diversity of tools helps train the learning algorithm and discriminate among the different tools themselves at run time. In the second step, a suitable

and recent set of features remaps and labels the empirical data. Finally, a machine-learning algorithm supports a two-layer classification model. A pair of specific classifiers implement the layers: one detects the category of the web attack that is being observed, while the second one recognizes the type of tool adopted. The latter layer operates on the top of the former classifier; it processes the detected attack's type and prompts the type of the software used to create the attack itself.

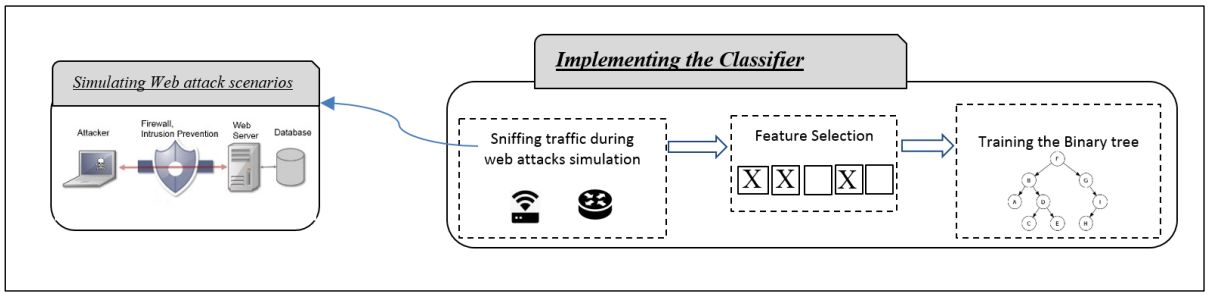

Figure 40. The work done to create the two-layers for web attacks classification

Figure 40 shows the three steps; it starts by dataset creation through simulating Web attacks. Then, the system applies the specific feature set to render those data into labeled (csv) files, each file corresponding to an attacking tool. Finally, a classification tree algorithm processes the resulting dataset and builds up the upper classification layer.

9.2.1 Dataset creation: Generating web attacks

In dataset creation phase, four types of web attacks were selected based on both the OWASP top 10 security risks and Positive Technologies 2019 report. These four web attacks are Brute Force, SQL Injection, XSS, and DOS. They are believed to be enough to reach our goal in analyzing the behavior of the attack generator, and in particular to show the ability of detecting the attacking tools used. Adding more webattacks might be part of future work. Moreover, the dataset we aim to create does not contain a benign traffic, as our main goal is to detect the tool used to generate a web attack, and though we consider that the classification is applied to already detected web attack. Generating the web attack was carried out against online testing

websites using both penetration tools and others used for attacking purposes. So for each attack type, a family of tools will be used. Moreover, to address the diversity of the behaviors of each attack, we used a set of tools to generate each web attack; these tools are used for penetration testing/ vulnerability assessment or for hacking purpose. Therefore:

1. To generate DOS, the following tools were used: Slowloris, R-U-Dead-Yet, Pyloris, Torshammer, LOIC.
2. To generate SQL Injection, the following tools were used: Fortify Webinspect, Acunetix, Jsql, Sqlmap, Arachni.
3. To generate XSS, the following tools were used: Fortify Webinspect, Acunetix, ZAP, Xsser, Arachni.
4. To generate Brute Force, the following tools were used**:** THC-Hydra, Medusa, Brutus, Wfuzz, Acunetix.

On the other hand, the websites used to simulate the attacks are online known ones used for penetration testing:

- http://zero.webappsecurity.com
- http://testphp.vulnweb.com
- http://testasp.vulnweb.com

The attacks represents thousands of HTTP requests sent by these specific tools. In parallel, the whole network traffic during the attack simulation was sniffed and saved in PCAP files, where each file corresponds to a specific tool. Therefore, each attacking tool will be represented by its own PCAP file containing the attacks it generated, and will be used to simulate its behavior. Files might be unbalanced in size, number of packets, duration of attack, and that is because of

the tool used, where each tool has its own way and configuration in which it acts or conducts the attack.

Details / Attack	Tool used	Attack duration (minutes)	Total Number of packets sniffed during the attack
Brute Force	*Acunetix*	10	115,000
	Brutus	19	138,000
	Hydra	6	28,000
	Medusa	4	33,000
	Wfuzz	14	50,000
Cross site scripting	*Acunetix*	31	35,000
	Arachni	30	60,000
	XSSER	2	10,600
	WebInspcet	7	27,000
Denial of service	*LOIC*	12	213,000
	Pyloris	15	249,000
	Rudy	32	79,000
	SlowLoris	5	269,000
	Torshammer	6	102,000
SQL Injection	*Acunetix*	More than 1 hour	13,000
	Arachni	50	5,000
	Jsql	25	8,000
	Sqlmap	More than 1 hour	6,000
	WebInspect	40	17,000

9.2.2 Feature selection and Labeling

In this work, a flow-based classification is used, such type of classification is referred when using properties (statistical features) such as flow bytes per second, duration per flow, etc. Using the statistical features of a specific IP traffic flow, we can get information about the nature of this flow, and might be applied to classify web attacks, and in particular to detect the attack generator. An IP traffic flow is by the 5-tuple composed of the following fields of the IP and TCP/UDP headers:

- IP source address
- IP destination address
- TCP/UDP source port
- TCP/UDP destination port
- Protocol

These fields are considered as 2-way (the inversion of source and destination ports and addresses is considered as one single flow).

However, the first task is to change the PCAP files containing the attacks into flow-based types selecting some basic components (features). A feature set containing 14 components (indicated as features) will be used to classify each flow are listed in the Table below. This feature set is created by [77], it was used mainly to detect whether a flow is affected by a malware or not, it is selected since it represents an effective feature set and it encompass the main characteristics of a network traffic, and also the use of these features is coherent with the literature in the field. In this chapter, we prefer concentrating on the main goal, and to leave refinements for further research. The experimental section will show how the following features associated to each flow are enough to infer the possible type of the web attack and software used. The features set contains: 'num_packets', 'tot_byte_flux', 'flow_duration', 'bit_rate', 'packet_rate',

'delta_mean', 'delta_std', 'DPL', 'first_len', 'max_len', 'min_len', 'mean_len', 'byte_std', 'data_entropy'. The next table describes each feature:

Table 14. Used features for each flow as statistical fingerprint

Features	Description
Num_Pack	Number of packets
Tot_Byte_Flux	Number of bytes
Flow_Duration	Duration of the flow in seconds
Byte_Rate	Byte rate
Packet_Rate	Packet rate
Delta_Mean	Average inter-arrival time of packets
Delta_Std	Standard deviation of inter-arrival time
LE	"Entropy" of the packet lengths
DPL	Total number of subsets of packets having the same length divided by the total number of packets of the flow
First_Len	Length of the first packet
Max_Len	Length of the longest packet
Min_Len	Length of the shortest packet
Mean_Len	Average packet length
Std_Len	Standard deviation of the packet length

Using this feature set, the IP network traffic of the PCAP files is converted into a CSV file containing group of flows, where the whole network traffic corresponding to the web attacks is changed into flows, and each flow is uniquely described by the vector of its features set. The vector of features represents the statistical fingerprint of the flow described in the table above.

Using this feature set, the network traffic will be changed into corresponding set flows of flows, the next table shows the number of flows for each web attack and each attacking tool.

Table 15. Changing attack sniffed traffic into the selected feature-set flows

Details / Attack	Tool used	Number of attack flows per tool	Total number of attack flows per attack type
Brute Force	Acunetix	11417	16616
	Brutus	1094	
	Hydra	2405	
	Medusa	1500	
	Wfuzz	200	
Cross site scripting	Acunetix	100	1942
	Arachni	1175	
	WebInspcet	107	
	XSSER	560	
Denial of service	LOIC	311	5081
	Pyloris	1000	
	Rudy	1420	
	SlowLoris	2100	
	Torshammer	250	
SQL Injection	Acunetix	748	2391
	Arachni	309	
	Jsql	565	
	Sqlmap	600	
	WebInspect	169	

After generating the flows, for SQL Injection, XSS, and brute force attacks, labeling was done by checking the packets of each flow, through inspecting the content of the HTTP requests,

making sure that the message sent on the application layer, contains an attack. DOS are labeled, based on the time when this attack was established, as all the connections where generating DOS attacks. Labeling the flows was done using the notation “Web Attack Type: Software type”, to identify each flow by its attack type and the attacking tool used to generate this attack.

9.2.3 Machine Learning Algorithm for Tool Recognition

Machine learning-based classifiers are aimed to identify to which set of categories a new sample belongs on the basis of a training set composed by data whose category is known. Machine learning has been used in several research areas to classify, find solutions, recognize patterns, etc… Machine learning tools form the basis of anomaly detection systems have proven to work with great success. In some cases, classifiers are used to discriminate normal from malicious traffic, in other cases to detect the type of the attack.

Literature proposes plenty of approaches based on the feature set and the eventual training algorithms. Most of the existing solutions are computationally demanding and implements classification strategies that cannot be interpreted by human users. In this study, we propose the use of binary classification trees [78]. This simple algorithm has a lot of advantages with respect to more sophisticated strategies: is one of the fastest approach in the literature, requires the storage of a very limited number of parameters and actually introduces only one hyperparameter, the number of levels. In addition, the computational cost of both the training and inference phase are quite limited. Finally, this algorithm is de-facto independent by data normalization [79]. This is a major advantage with respect to other algorithms that suffers of numerical instability. In addition, the classification procedure is based on a set of thresholds. As a major consequence, the prediction process can be the rules employed by the classifier can be evaluated by data analyst.

This chapter proposes a two-layers classification model to detect the type of the attacking tool. Figure 5 shows how the two classification layers are trained independently. Once the final models are trained, the incoming data are firstly classified by the "Attack type" classifier. Later, for each class of attacks, a specialized classifier "Tool type" assigns the incoming pattern to the generating tool. Again should be noted that in the proposed study both attack classifier and tool type classifier are implemented by means of classification tree. As a consequence the decision rule can be understand by data analysts.

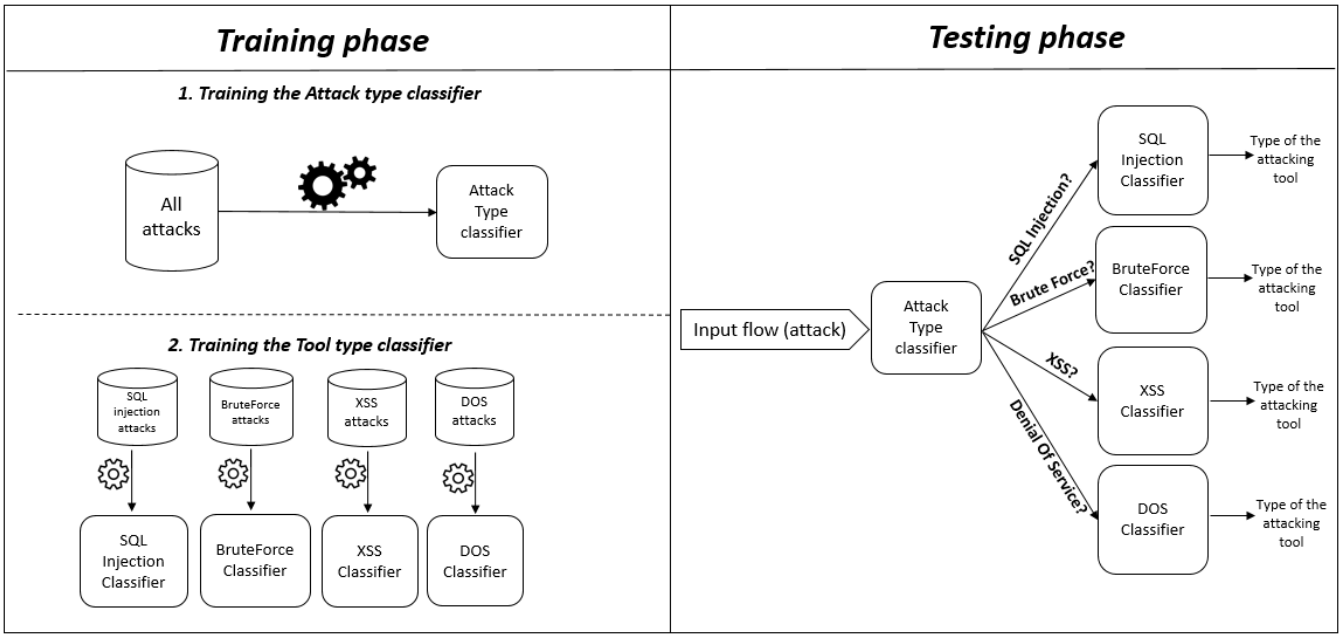

Figure 41. Training and testing the two-layers classification for webattacks

9.3 Experimental section:

The evaluation of the method's effectiveness at classifying attack-generation tools covered two steps. First, a classifier discriminated the four categories of attacks. This phase is critical in a real world scenario, where no prior information is available about the type of the web attack. Next, the data split for each category are analyzed independently to recognize the attack-generating tool.

9.3.1 First classification layer: Attacks Classification

A tree with a layered hierarchy of threshold-based classifiers supported the recognition of the attack types. The dataset was randomly split into a training set including a portion of 70% of the available samples, and a test set holding the remaining 30%. To ensure experiment repeatability, the parameters were all set to the default values of the scikit-learn python toolbox. Table 16 reports on the classification performances measured on the test set, i.e., by classifying patterns that had not been used for training. The confusion matrix gives, in each column, the actual predictions of the proposed model, and, in each row, the actual labels. Ideally, a perfect classifier would attain a diagonal matrix in which predictions and true labels always coincide. The results reported refer to a classifier with a maximum depth equal to five, which means that the final prediction is obtained controlling five thresholds.

Table 16. Confusion matrix testing the first classification layer: classifier of Web attack type with max depth =5

Confusion matrix for the first classification layer		**Classified as:**			
		BF	**XSS**	**DOS**	**SQL injection**
Testing attacks	**BF flows**	*5496*	*8*	*0*	*0*
	XSS flows	*1*	*594*	*0*	*23*
	DOS flows	*0*	*0*	*1722*	*0*
	SQL injection flows	*3*	*26*	*1*	*768*

Empirical evidence confirmed that a simple classification rule could anyway yield satisfactory performances.

Table 17 gives the results obtained when using a state-of-the-art classifier model in terms of accuracy, i.e., a random forest classifier, which embeds an ensemble of classification trees. The number of classification trees was set to 100 by using a subset of training data as a validation set. The drawback of any random forest classifier, however, is that it implements a sort of

black-box model, as human operators cannot inspect or interpret the actual decision strategy easily. Moreover, the computational cost and the number of parameters of Random-Forest models grow linearly with the number of trees, hence in this experimental setup the cost was 100 times larger than that associated with the basic classification tree reported in Table 16.

Table 17. Confusion matrix testing the first classification layer: classifier of Web attack Attack-type using a baseline random forest classifier with n_tree=100

Confusion matrix for the first classification layer		**Classified as:**			
		BF	**XSS**	**DOS**	**SQL injection**
Testing attacks	**BF flows**	*5504*	*0*	*0*	*0*
	XSS flows	*0*	*613*	*0*	*5*
	DOS flows	*0*	*0*	*1721*	*1*
	SQL injection flows	*1*	*24*	*2*	*771*

The experimental outcomes confirmed that Random Forest did provide a considerable improvement in the classification of XSS patterns, as the overall number of errors decreased from 23 to 5. At the same time, the difference proved minimal in the other cases.

In summary, this experimental campaign confirmed that fast and white-box ML algorithms can profitably support Web Attacks classification.

9.3.2 Second classification layer: Tool Classification

The upper classification layer aimed to identify the type of software, and a specific classifier was built for each flow associated with a category of web attack. In other words, the preliminary classification performed by the lower layer determines the type of attack, and for each type of attack a specific classifier is built up to identify the generating tool. Therefore, the research presented here involved four classifiers.

For each attack category, a training set again include a fraction of 70 % of the available attack; whereas the remaining data (30%) formed the test set.

The number of levels in the classification tree was set to 10 as in the previous experimental section. Tables 18,19,20,21 report on the measured accuracy in tool classification for the attack types SQL Injection, brute force, XSS, and DOS, respectively. The tables give results in the form of confusion matrix.

Table 18. Confusion matrix testing the classifier used to detect the tool that generate SQL injection attacks

The Confusion matrix		Classified as				
		Acunetix	**Arachni**	**JSQL**	**SQLMap**	**WebInspect**
Testing attacks generated by	**Acunetix**	257	0	0	0	3
	Arachni	0	95	0	1	0
	JSQL	0	0	183	0	1
	SQLMap	0	0	2	198	0
	WebInspect	1	0	0	0	47

Table 19. Confusion matrix testing the classifier used to detect the tool used to generate BRUTE FORCE attacks:

The Confusion matrix		Classified as				
		Acunetix	**Brutus**	**Hydra**	**Medusa**	**Wfuzz**
Testing attacks generated by	**Acunetix**	3802	0	0	0	0
	Brutus	0	348	0	0	0
	Hydra	0	0	796	0	0
	Medusa	0	0	1	499	0
	Wfuzz	0	0	0	0	61

Table 20 Confusion matrix testing the classifier used to detect the tool that generates XSS attacks:

The Confusion matrix		Classified as			
		Acunetix	**Arachni**	**WebInspet**	**XSSER**
Testing attacks generated by	**Acunetix**	30	1	0	0
	Arachni	0	377	0	0
	WebInspet	0	0	36	3
	XSSER	0	0	0	197

Table 21. Confusion matrix testing the classifier used to detect the tool that generates DOS attacks

The Confusion matrix		Classified as				
		LOIC	**Pyloris**	**Rudy**	**Slowloris**	**Torshammer**
Testing attacks generated by	**LOIC**	116	0	0	0	0
	Pyloris	0	337	0	0	0
	Rudy	1	0	467	0	0
	Slowloris	0	0	0	706	0
	Torshammer	0	0	0	0	78

9.4 Result analysis:

When assembling the test results for both the lower-level classifier (attack-type classifier, as per Table 17), and the upper level classifier tree (Attacking tool classifier, as per Tables 18-21), one verifies the effectiveness of the proposed method to recognize both the attack's type and the tool used to generate it. More importantly, the proposed approach stands on a very efficient yet interpretable model. As a consequence, analysts and engineers can benefit from the resulting approach to evaluate the statistical properties of advanced attacking tool.

The chapter considered a collection of the most popular categories of web attacks, generated using a variety of tools. The statistical properties of the gathered data fed white-box machine-learning models, allowing easy inspection and interpretation by human supervisors. Empirical results confirmed that the set of statistical features could support the discrimination of different attack schemes and generating tools. In the future, we aim to increase the number of web attacks and to create a public repository, available to the research community, to enrich a public dataset on a daily basis, through adding web attacks sniffed at the network level.

10 Conclusions and future works

Data must be treated as the core asset of any organization or company, which should be protected from all kinds of threats. However, to best protect the data, it is better to adopt a good management approach, which minimizes errors, saves time, increases awareness and prepares the company to incidents. In information security, taking due care of strategies, setting policies, plans, preparing and training staff, and complying to internal or external laws and standards should be given the same importance as operating technical tools. The focus should be on both organizational and technical sides. Organizations in general including companies should gives a great significance to the organizational side of data protection, which is shown by its ISMS which was created in accordance with the international standards and frameworks. They should also verify that the ISMS is working effectively in the sense that all the defined requirements are correctly implemented, and this can be obtained by regular audits to reach continuous improvement. This PhD thesis handled the issue of managing data protection, aiming to protect the company's data from both the organizational and the technological side, by implementing an Information Security Management System (ISMS), an ISMS consists of the policies, procedures, guidelines, and associated resources and activities, collectively managed by an organization, in the pursuit of protecting its information asset. The goal of ISMS is to minimize risk and ensure continuity by pro-actively limiting the impact of a security breach. After that, a comprehensive framework is designed for the security risk management of Cyber physical systems, this framework represents the strategy used to manage the security risk management, and it falls inside the ISMS as a security strategy. In addition, inside the ISMS, another part of the work suggested an approach to select an applicable Security Incident and Event Management (SIEM) solution, to support companies that are seeking to adopt SIEM systems. This approach, unlike others, is customer driven which means that customer needs are taken

into account when following the whole approach, specifically when defining the requirements and then evaluating the suppliers' solutions.

Hitachi Rail STS realizes the necessity to verify that the ISMS is working effectively in the sense that all the defined requirements are correctly implemented, and this can be obtained by regular audits to reach continuous improvement. Hitachi Rail STS also realizes that not all assets are correctly managed, in fact, there are some areas not completely covered such as laboratories, plants connections and so on. Therefore, in the future, it is important to extend the coverage of the ISMS to these areas and not only to office areas. Also Hitachi Rail STS is going to comply with the international standard ISO 27001 on some strategic scopes with the aim to obtain a certification, which represents a key goal for the company and its business.

At the end, a research topic was carried out to detect the type of the webattacks and the tool used to generate a web attack. Therefore, a two-layers classification system was designed using network statistical fingerprints and machine learning techniques to detect the type of the software used by the attackers to generate web attacks, since any information about the attackers might be helpful and worth a lot to the cyber security analysts and might aid in the prevention and mitigation of web attacks.

In conclusion, this thesis handled more the one topic that aims to protect data that all falls inside the ISMS that has a main goal, which is managing data protection. It deals more with the organizational side of data protection, but also it presented at the end an important contribution "A two layers classification model" to classify the type and the tool of a web attack, which can be placed in the technical side of a risk management strategy and aids in the mitigation of web attacks.

11 References

[1] A. S. Elmaghraby and M. Losavio, "Cyber security challenges in smart cities: Safety, security and privacy," *Journal of Advanced Research,* 2014.

[2] I. Internet Systems Consortium, "Internet Domain Survey," January 2019. [Online]. Available: https://www.isc.org/survey/.

[3] S. Morgan, "Herjavec Group: A Guide To Cybersecurity Conversations For The C-Suite in 2019," 4 August 2019. [Online]. Available: https://www.herjavecgroup.com/cybersecurity-conversations-2019/.

[4] S. Morgan, "Cybercrime Damages $6 Trillion By 2021," 7 December 2018. [Online]. Available: https://cybersecurityventures.com/cybercrime-damages-6-trillion-by-2021/.

[5] S. C. Bennett, "Security Breaches: Problems And Solutions," The Practical Lawyer, 2008.

[6] A. Arora, A. Nandkumar and R. Telang, "Does information security attack frequency increase with vulnerability disclosure? An empirical analysis," in *Proceedings of the International Conference on Information Systems, ICIS 2006*, Milwaukee, Wisconsin, USA, 2006.

[7] PECB, "The importance of information security nowadays," [Online]. Available: https://pecb.com/pdf/articles/27-pecb_the-importance-of-information-security-nowadays.pdf.

[8] H. Mokalled, D. Debertol, E. Meda and C. Pragliola, "The importance to manage data protection in the right way: Problems and solutions," in *Optimization and decision science: methodologies and applications: ODS.*, Sorrento, Italy, Springer, September 2017, p. 69–82..

[9] H. Mokalled, C. Pragliola, D. Debertol, E. Meda and R. Zunino, "A Comprehensive Framework for the Security Risk Management of Cyber-Physical System," in *Resilience of Cyber-Physical Systems. Advanced Sciences and Technologies for Security Applications.*, Springer, 2019, pp. 49-68.

[10] H. Mokalled, R. Catelli, V. Casola, D. Debertol, E. Meda and R. Zunino, "The Applicability of a SIEM Solution: Requirements and Evaluation," in *2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, Napoli, Italy, 2019.

[11] B. S. and A. S., "A Roadmap to Data Security of Automated University Examination System," *International Journal of Scientific and Technical Advancements,* pp. 71-75, 2015.

[12] G. Summers, "Data and databases," in *eveloping Databases with Access*, Nelson Australia Pty Limited, 2004, pp. 4-5.

[13] "Annual Emerging Cyber Threats Report," 2013. [Online]. Available: https://cyber.gatech.edu/threats-reports.

[14] Symantec, "Internet Security Threats Report," 2013. [Online]. Available: http://www.symantec.com/threatreport/.

[15] D. Cappelli, A. Moore and R. Trzeciak, "The CERT guide to insider threats: How to prevent, detect, and respond to theft of critical information, sabotage, and fraud," 2012. [Online]. Available: http://ptgmedia.pearsoncmg.com/images/9780321812575/samplepages/9780321812575.pdf.

[16] J. Hunker and C. W. Probst, "Insiders and Insider Threats- An Overview of Definitions and Mitigation Techniques," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications,* pp. 4-27, 2011.

[17] A. F., E. P., P. M., C. H. and F. S., "Decision support approaches for cyber security investment," *Decision Support Systems,* pp. 13-23, 2016.

[18] "International standard ISO_IEC_27000," ISO, 2014.

[19] "International standard ISO_27001," 2013.

[20] Y. Peng, T. Lu, J. Liu, Y. Gao, X. Guo and F. Xie, "Cyber-Physical System Risk Assessment," in *Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2013.

[21] "CBTC Communication Based Train Control," [Online]. Available: http://sts.hitachirail.com/sites/ansaldosts/files/imce/asts_hitachi_cbtc_ingl_lr.pdf.

[22] Siemens, "Trainguard sirius CBTC," 2013. [Online]. Available: http://www.mobility.siemens.com/mobility/global/SiteCollectionDocuments/en/rail-solutions/rail-automation/train-control-systems/trainguard-sirius-cbtc-en.pdf.

[23] B. Chen, C. Schmittner, Z. Ma, W. G. Temple, X. Dong, D. L. Jones and W. H. Sanders, "Security Analysis of Urban Railway Systems: The Need for a Cyber-Physical Perspective," in *International Conference on Computer Safety, Reliability, and Security*, 2015.

[24] "Hitachi Rail Signalling and Transportation Systems (Hitachi Rail STS)," 2019. [Online]. Available: http://sts.hitachirail.com/en.

[25] Enisa, "MEHARI- ENISA," ENISA, [Online]. Available: https://www.enisa.europa.eu/topics/threat-risk-management/risk-management/current-

risk/risk-management-inventory/rm-ra-methods/m_mehari.html. [Accessed February 2020].

[26] Enisa, "EBIOS- Enisa," ENISA, [Online]. Available: https://www.enisa.europa.eu/topics/threat-risk-management/risk-management/current-risk/risk-management-inventory/rm-ra-methods/m_ebios.html. [Accessed February 2020].

[27] ENISA, "ENISA Threat Landscape Report 2017, 15 Top Cyber-Threats and Trends," 2017. [Online]. Available: https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-2017.

[28] MAGERIT, "MAGERIT – version 3.0 Methodology for Information Systems Risk Analysis and Management, Book I - The Method," MAGERIT, Madrid, 2014.

[29] PILAR, "PILAR Risk Analysis and Management, Help Files version 6.2," PILAR, 2016.

[30] V. Casola, A. Fasolino, N. Mazzocca and P. Tramontana, "An AHP-based framework for quality and security evaluation," in *12th IEEE International Conference on Computational Science and Engineering*, 2009.

[31] N. Miloslavskaya, "Analysis of SIEM Systems and Their Usage in Security Operations and Security Intelligence Centers.," in *Biologically Inspired Cognitive Architectures (BICA) for Young Scientists.*, Springer, 2018.

[32] Verizon, "Verizon in the Data Breach Investigations Report," Verizon , 2015.

[33] IBM, "IT Security Compliance Management Design Guide with IBM Tivoli Security Information and Event Manager. 2nd edn.," 2010. [Online]. Available: http://www.redbooks.ibm.com/abstracts/sg247530.html?Open.

[34] G. Sadowski, K. Kavanagh and T. Bussa, "Technology Insight for the modern SIEM," Gartner, 2018.

[35] T. Bussa, K. Kavanagh and G. Sadowski, "Critical Capabilities for security Information and Event management," Gartner, 2018.

[36] Techtarget, "Techtarget: Security information and event management (SIEM)," 2014. [Online]. Available: http://searchsecurity.techtarget.com/definition/security-information-and-event-management-SIEM..

[37] K. Scarfone, "Introduction to SIEM services and products," 2015. [Online]. Available: http://searchsecurity.techtarget.com/feature/Introduction-to-SIEM-services-and-products.

[38] K. Kavanagh, T. Bussa and G. Sadowski, "Magic Quadrant for Security Information and Event Management," Gartner, 2018.

[39] O. Rochford, K. Kavanagh and T. Bussa, "Critical Capabilities for Security Information and Event Management," Gartner, 2016.

[40] SANS, "Benchmarking Security Information Event Management (SIEM)," SANS Institute InfoSec Reading Room, 2009.

[41] M. Nabil, S. Soukainat, A. Lakbabi and O. Ghizlane, "SIEM selection criteria for an efficient contextual security," in *International Symposium on Networks, Computers and Communications (ISNCC)*, Marrakech, 2017.

[42] M. Scriven, "The methodology of evaluation," in *Stake, R. E. Curriculum evaluation*, Chicago, American Educational Research Association, 1967.

[43] T. Banta and C. Palomba, Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education, San Francisco: Jossey-Bass, Inc., 1999.

[44] A. Elmaghraby and M. Losavio, "Cyber Security Challenges in Smart Cities: Safety, security and privacy.," *Journal of Advanced Research,* 2014.

[45] InternetLiveStats, "Total number of Websites- Internet Live stats," InternetLiveStats, 10 November 2019. [Online]. Available: https://www.internetlivestats.com/total-number-of-websites/#ref-1. [Accessed 10 November 2019].

[46] Netcraft, "How many active sites are there?," Netcraft, November 2019. [Online]. Available: https://www.netcraft.com/active-sites/. [Accessed 10 November 2019].

[47] "SiteLock- About us," SiteLock, [Online]. Available: https://www.sitelock.com/about/. [Accessed 10 Novermber 2019].

[48] SiteLock, "The Secret Life of Websites: SitLock website security Insider [Q1 2018]," 2018. [Online]. Available: https://www.sitelock.com/download/SiteLock%20Website%20Security%20Insider%20Q1%202018.pdf. [Accessed 10 November 2019].

[49] I. Sharafaldin, A. Habibi Lashkari and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," in *4th International Conference on Information Systems Security and Privacy (ICISSP)*, Portugal, 2018.

[50] OWASP, "OWASP Top Ten Project," OWASP, 2017. [Online]. Available: https://www.owasp.org/images/7/72/OWASP_Top_10-2017_%28en%29.pdf.pdf. [Accessed 10 November 2019].

[51] PositiveTechnologies, "Positive Technologies- About us," Postive Technologies, 2019. [Online]. Available: https://www.ptsecurity.com/ww-en/about/. [Accessed 10 November 2019].

[52] PositiveTechnologies, "Attacks on web applications: 2018 in review: Web-application-attacks-2019-Eng," Positive Technologies, 2018. [Online]. Available: https://www.ptsecurity.com/upload/corporate/ww-en/analytics/Web-application-attacks-2019-eng.pdf. [Accessed 10 November 2019].

[53] A. Buczak and G. Erhan, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," *IEEE Communications Surveys & Tutorials,* vol. 18, no. 2, pp. 1153 - 1176, 2015.

[54] S. Mukherjee and N. Sharma, "Intrusion Detection using Naive Bayes Classifier with Feature Reduction," *Procedia Technology,* p. 119–128, 2012.

[55] J. Li, Y. Qu, F. Chao, H. P. H. Shum, E. S. L. Ho and L. Yang, "Machine Learning Algorithms for Network Intrusion Detection," in *AI in Cybersecurity*, Springer, Cham, 2018, pp. 151-179.

[56] Z. Ma and A. Kaban, "K-Nearest-Neighbours with a novel similarity measure for intrusion detection," in *2013 13th UK Workshop on Computational Intelligence (UKCI)*, Guildford, 2013.

[57] A. H. Sung and S. Mukkamala, "Feature Selection for Intrusion Detection with Neural Networks and Support Vector Machines," *Journal of the Transportation Research board,* vol. 1822, no. 1, 2003.

[58] M. Kumar, M. Hanumanthappa and K. TVS, "Intrusion detection system using decision," in *Proceedings of the 14th IEEE International Conference on Communication*, New York, 2012.

[59] N. Farnaaz and M. Jabbar, "Random Forest Modeling for Network Intrusion Detection System," *Procedia Computer Science,* vol. 89, pp. 213-217, 2016.

[60] U. Ravale, N. Marathe and P. Padiya, "Feature Selection Based Hybrid Anomaly Intrusion Detection System Using K Means and RBF Kernel Function," *Procedia Computer Science,* vol. 45, pp. 428-435, 2015.

[61] H. G. Kayacik, A. N. Zincir-Heywood and M. I. Heywood, "A hierarchical SOM-based intrusion detection system," *Engineering Applications of Artificial Intelligence,* vol. 20, no. 4, pp. 439-451, 2007.

[62] R. Lippmann, R. K. Cunningham, D. J. Fried, I. Graf, K. R. Kendall, S. E. Webster and M. A. Zissman, "Results of the DARPA 1998 Offline Intrusion Detection Evaluation," in *Recent Advances in Intrusion Detection, RAID 99 Conference*, Indiana, USA, 1999.

[63] R. Lippmann, J. W. Haines, D. J. Fried, J. Korba and K. Das, "The 1999 DARPA off-line intrusion detection evaluation," *Computer Networks,* vol. 34, no. 4, pp. 579-595, 2000.

[64] "KDD Cup 1999 Data," 28 October 1999. [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html. [Accessed 11 Novermber 2019].

[65] J. McHugh, "Testing Intrusion Detection Systems: A Critique of the 1998 and 1999 DARPA Intrusion Detection System Evaluations as Performed by Lincoln Laboratory," *ACM Transactions on Information and System Security,* vol. 3, no. 4, p. 262– 294, Novermber 2000.

[66] C. Brown, A. Cowperthwaite, A. Hijazi and A. Somayaji, "Analysis of the 1999 DARPA/Lincoln Laboratory IDS evaluation data with NetADHICT," in *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications,* , Ottawa, ON, Canada, July 2009.

[67] M. Tavallaee, E. Bagheri, W. Lu and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, Ottawa, ON, Canada, July 2009.

[68] "The CAIDA OC48 Peering Point Traces," Center for Applied Internet Data Analysis, 2002. [Online]. Available: https://www.caida.org/data/passive/passive_oc48_dataset.xml. [Accessed 11 November 2019].

[69] "The CAIDA "DDoS Attack 2007" Dataset," Center for Applied Internet Data Analysis, 2007. [Online]. Available: https://www.caida.org/data/passive/ddos-20070804_dataset.xml. [Accessed 11 November 2019].

[70] "The CAIDA Anonymized Internet Traces 2016 Dataset," Center for Applied Internet Data Analysis, 2016. [Online]. Available: https://www.caida.org/data/passive/passive_2016_dataset.xml. [Accessed 11 November 2019].

[71] E. P. Proebstel, "Characterizing and improving distributed networkbased intrusion detection systems(nids): timestamp," Master's thesis, University of California DAVIS, CA, USA, 2008.

[72] E. Sangster, T. J. O'Connor, T. Cook, R. Fanelli, E. Dean, W. J. Adams, C. Morrell and G. Conti, "Toward instrumenting network warfare," United States Military Academy, New York, 2007.

[73] A. Habibi Lashkari, G. Draper Gil, M. Saiful Islam Mamun and A. A. Ghorbani, "Characterization of Tor Traffic using Time based Features," in *3rd International Conference on Information Systems Security and Privacy*, 2017.

[74] M. Aiello, M. Mongelli and G. Papaleo, "DNS tunneling detection through statistical fingerprints of protocol messages and machine learning," *International Journal of Communication Systems,* vol. 28, no. 14, July 2014.

[75] M. Dusi, M. Crotti, F. Gringoli and L. Salgarelli, "Tunnel Hunter: Detecting application-layer tunnels with statistical fingerprinting," *Computer Networks,* vol. 53, no. 1, pp. 81-97, 2009.

[76] A. Gharib, I. Sharafaldin, A. H. Lashkari and A. A. Ghorbani, "An Evaluation Framework for Intrusion Detection Dataset," in *International Conference on Information Science and Security (ICISS)*, Pattaya, 2016.

[77] L. Boero, M. Cello, M. Marchese, E. Mariconti, T. Naqash and S. Zappatore, "Statistical fingerprint-based intrusion detection system (SF-IDS)," *International Journal of Communication Systems,* October 2016.

[78] W. Buntine, "Learning Classification Trees," *Statistics and Computer journal,* vol. 2, no. 2, pp. 63-73, 1992.

[79] L. Breiman, "Random Forests," *Machine Learning,* vol. 45, no. 1, pp. 5-32, October 2001.

[80] G. Biau and E. Scornet, "A random forest guided tour," *Test,* vol. 25, no. 2, pp. 197-227, June 2016.